MTH-RA, Real Analysis

Atonu Roy Chowdhury,
achowdhu@ictp.it, atonuroychowdhury@gmail.com

Last updated: 16th November, 2026, 3:20pm

General Information About the Course

- **Instructor:** Emanuel Carneiro

  Email: carneiro@ictp.it

  ICTP Webpage: https://www.ictp.it/member/emanuel-carneiro

  Personal website: https://sites.google.com/view/emanuelcarneiro

  Office: LB-109

No AI (or LLM) has been used to preapre these notes. All errors are my own. Please send me an email if anything is wrong, or not clearly explained.

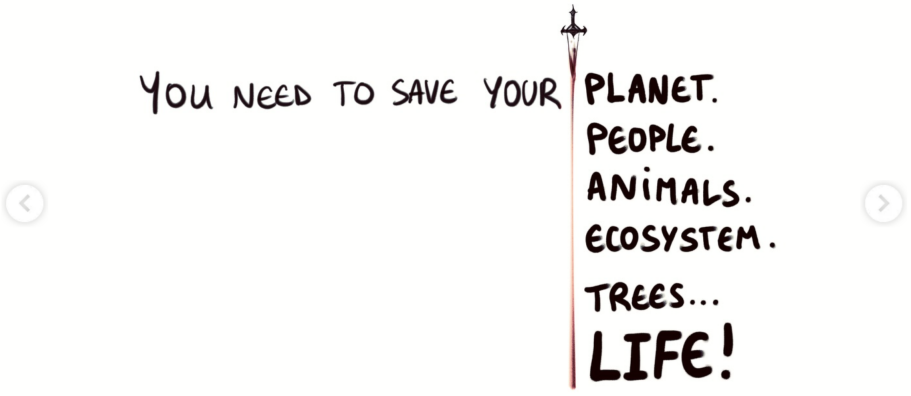


Figure 1: Original artist: lotycaa on Instagram. Source.

Contents

1. 10 September 2026 **6**

1.i. σ-algebras and examples . . . 6
1.ii. Measures . . . 9
1.iii. Finite and σ-finite measures . . . 10
1.iv. Measurable Functions . . . 12
1.v. Basic properties of measurable functions . . . 16
1.vi. Complete measure . . . 17
1.vii. Exercises . . . 20

2. 11 September 2026 **22**

2.i. Motivation and Characteristic Functions . . . 22
2.ii. Simple Functions . . . 22
2.iii. Approximation of Non-negative Functions by Simple Functions . . . 26
2.iv. The Integral of a Non-negative Measurable Function . . . 27
2.v. Basic Properties of the Integral . . . 27
2.vi. The Monotone Convergence Theorem . . . 28

3. 14 September 2026 **35**

3.i. Fatou's lemma . . . 35
3.ii. Integrable functions . . . 37
3.iii. Dominated convergence theorem . . . 40
3.iv. A non-measurable set . . . 43

4. 16 September 2026 **46**

4.i. Additivity of Volume over Almost Disjoint Partitions . . . 46
4.ii. Subadditivity of Volume over Coverings . . . 47
4.iii. The Structure of Open Sets in $\mathbb{R}$ . . . 49
4.iv. The Structure of Open Sets in $\mathbb{R}^d$, $d \geq 2$ . . . 50

5. 17 September 2026 **53**

5.i. Exterior Measure . . . 53
5.ii. Basic Properties of Exterior Measure . . . 54
5.iii. Lebesgue Measurable Sets . . . 60

6. 18 September 2026 **64**

6.i. More properties of Lebesgue measure . . . 64
6.ii. Completeness of Lebesgue measure . . . 68
6.iii. Invariance properties of Lebesgue measure . . . 70
6.iv. Lebesgue measure and linear maps . . . 71
6.v. Exercises . . . 78

7. Extra material: General outer measure **90**
7.i. Outer measures . . . 90
7.ii. Carathéodory's theorem . . . 91
7.iii. Pre-measures . . . 93
7.iv. Exercises . . . 96
7.v. Borel measures on $\mathbb{R}$ . . . 101

8. 21 September 2026 **106**
8.i. Egorov's theorem . . . 106
8.ii. Lusin's theorem . . . 107
8.iii. Exercises . . . 110

9. 23 September 2026 **112**
9.i. L^p spaces . . . 112
9.ii. Basic Properties of the L^p spaces . . . 112
9.iii. Young's Inequality . . . 113
9.iv. Hölder's Inequality . . . 114
9.v. Minkowski's Inequality . . . 118

10. 24 September 2026 **121**
10.i. The Riesz–Fischer Theorem . . . 121
10.ii. Density properties of L^p spaces . . . 125
10.iii. Exercises . . . 131

11. 28 September 2026 **159**
11.i. Duality of the norm . . . 159
11.ii. Distribution function and weak L^p . . . 162
11.iii. Integrability of powers, and how to test $f \in L^p$ . . . 163
11.iv. The layer-cake formula . . . 164
11.v. Jensen's inequality . . . 164

12. 30 September 2026 **168**
12.i. Inner products and norms . . . 168
12.ii. Examples . . . 170
12.iii. Closed subspaces and orthogonal complements . . . 171
12.iv. The projection theorem . . . 172

13. 1 October 2026 **175**
13.i. When are L^p spaces Hilbert spaces . . . 175
13.ii. Bounded linear functionals and the dual space . . . 177
13.iii. The Riesz representation theorem . . . 180
13.iv. Signed measures . . . 181
13.v. The total variation measure . . . 182
13.vi. The space of signed measures . . . 185

14. 12 October 2026 **188**
14.i. Absolute continuity and singularity . . . 188
14.ii. A finite equivalent of a σ-finite measure . . . 191
14.iii. The Lebesgue–Radon–Nikodym theorem . . . 192
14.iv. Bounded functionals in L^p: Riesz representation revisited . . . 198
14.v. Exercises . . . 201

15. 14 October 2026 **208**
15.i. Fubini's theorem . . . 209
15.ii. Tonelli's theorem . . . 214
15.iii. The case of general measures . . . 219
15.iv. Exercises . . . 223

16. 15 October 2026 **233**
16.i. The fundamental theorem of calculus . . . 233
16.ii. Hardy–Littlewood maximal function . . . 234
16.iii. Weak-type $(1,1)$ estimate implies Lebesgue differentiation theorem . . . 238
16.iv. The Vitali covering lemma . . . 240
16.v. Proof of the weak-type $(1,1)$ estimate . . . 240
16.vi. Exercises . . . 243

17. 29 October 2026 **249**
17.i. Functions of bounded variation . . . 249
17.ii. Dini numbers and the main theorem . . . 250
17.iii. A second Vitali covering lemma . . . 251
17.iv. Proof of differentiability of BV functions . . . 253

18. 4 November 2026 **258**
18.i. Absolutely continuous functions . . . 258
18.ii. Fundamental theorem of calculus for AC functions . . . 259
18.iii. Exercises . . . 261

19. 5 November 2026 **267**
19.i. The convolution operator . . . 267
19.ii. The regularizing effect . . . 268
19.iii. Convolution of L^p functions . . . 269
19.iv. Young's inequality . . . 271
19.v. Approximations of the identity . . . 274

20. 6 November 2026 **278**
20.i. Pointwise convergence of approximations of the identity . . . 278
20.ii. Reduction to the Hardy–Littlewood maximal function . . . 280
20.iii. Proof of Approximations of the identity: pointwise convergence . . . 282
20.iv. The Fourier transform . . . 285

20.v. Basic properties of Fourier transform 286
20.vi. Exercises 290

21. 9 November 2026 **309**
21.i. Recap: duality between decay and regularity 309
21.ii. Examples 310
21.iii. The Riemann–Lebesgue lemma 313
21.iv. The multiplication formula and the inversion theorem 314
21.v. The L^2 theory 317
21.vi. The Fourier transform in L^p for $1 \leq p \leq 2$ 321
21.vii. The Schwartz class 324
21.viii. Exercises 325

References **349**

§1. 10 September 2026

§1.i. σ-algebras and examples

The classical motivation for measure theory goes back more than two thousand years: the Greeks wanted to measure the area of a circle, the volume of a ball, and other regions in the plane and in space. This eventually evolved into an abstract theory: instead of working only in $\mathbb{R}^d$, we work with an arbitrary set Ω and ask how to assign a consistent notion of "size" to certain of its subsets.

A key obstruction appears immediately:

- It is **not possible**, in general, to consistently measure *every* subset of a set (even of $\mathbb{R}^d$).
- Instead, we single out a suitable collection of "good" subsets (we call them the **measurable sets**) and only attempt to measure those.

This leads to the first fundamental definition of the course: the notion of a σ-algebra.

Definition 1.1 (σ-algebra)**.** Let Ω be a set. A collection Σ of subsets of Ω is called a **σ-algebra** (or σ-field) on Ω if:

(i) $\Sigma \neq \varnothing$ (equivalently, $\Omega \in \Sigma$);

(ii) (closed under complements) if $A \in \Sigma$, then $A^c := \Omega \setminus A \in \Sigma$;

(iii) (closed under countable unions) if $A_1, A_2, A_3, \ldots \in \Sigma$, then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

The pair (Ω, Σ) is called a **measurable space**, and the sets $A \in \Sigma$ are called **measurable sets**.

Remark 1.1. Condition (i) simply rules out the degenerate empty collection, which would vacuously satisfy (ii) and (iii) but is of no interest. Some textbooks instead require $\varnothing \in \Sigma$ directly; this is equivalent given (ii)–(iii) and a nonempty Σ.

Proposition 1.1

Every σ-algebra Σ on Ω satisfies:

(a) $\varnothing \in \Sigma$ and $\Omega \in \Sigma$;

(b) Σ is closed under countable intersections: if $A_1, A_2, \ldots \in \Sigma$ then $\bigcap_{n=1}^{\infty} A_n \in \Sigma$;

(c) Σ is closed under finite unions and finite intersections (in particular under set differences).

Proof. (a) Since $\Sigma \neq \varnothing$, pick $A \in \Sigma$. By (ii), $A^c \in \Sigma$, and by (iii) (a union of the two, or of A and infinitely many copies of it), $A \cup A^c = \Omega \in \Sigma$. Applying (ii) again, $\Omega^c = \varnothing \in \Sigma$.

(b) By De Morgan's law,

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c \right)^c . \tag{1.1}$$

Each $A_n^c \in \Sigma$ by (ii), the union is in Σ by (iii), and its complement is in Σ by (ii) again.

(c) A finite union $A_1 \cup \cdots \cup A_k$ is a countable union with $A_n = \varnothing$ for $n > k$ (using $\varnothing \in \Sigma$ from (a)), so it lies in Σ by (iii). Finite intersections follow similarly from (b), and $A \setminus B = A \cap B^c \in \Sigma$.

■

Remark 1.2. The definition does *not* require the sets A_n in condition (iii) to be pairwise disjoint; the countable union of *any* sequence of measurable sets is measurable.

Example 1.1 (Trivial and power set σ-algebras)

For any Ω:

- $\{\varnothing, \Omega\}$ is the smallest possible σ-algebra on Ω.
- The power set $\mathcal{P}(\Omega)$ (all subsets of Ω) is the largest possible σ-algebra on Ω; if $|\Omega| = n < \infty$ then $|\mathcal{P}(\Omega)| = 2^n$.

Example 1.2 (Generated σ-algebra)

Let X be an arbitrary collection of subsets of Ω. There is a **smallest** σ-algebra $\sigma(X)$ containing X, called the σ-algebra **generated by** X. It can be constructed as

$$\sigma(X) = \bigcap \{\Sigma : \Sigma \text{ is a } \sigma\text{-algebra on } \Omega,\ X \subseteq \Sigma\}, \tag{1.2}$$

i.e. the intersection of *all* σ-algebras containing X (an arbitrary intersection of σ-algebras is again a σ-algebra, and $\mathcal{P}(\Omega)$ is one such σ-algebra containing X, so this intersection is nonempty and well defined).

Lemma 1.2

Let Ω be a set and $\{\Sigma_\alpha\}_{\alpha \in J}$ an arbitrary collection of σ-algebras on Ω. Then $\bigcap_{\alpha \in J} \Sigma_\alpha$ is also a σ-algebra on Ω.

Proof. Every Σ_α contains $\varnothing$ and Ω, so $\Sigma := \bigcap_{\alpha \in J} \Sigma_\alpha$ is nonempty.

Suppose $A \in \Sigma$. Then $A \in \Sigma_\alpha$ for every α. Since Σ_α is a σ-algebra, it contains A^c, for every α. Therefore, $\Sigma := \bigcap_{\alpha \in J} \Sigma_\alpha$ also contains A^c.

Suppose $A_n \in \Sigma$ for $n \in \mathbb{N}$. Then $A_n \in \Sigma_\alpha$ for every α, for every n. Since Σ_α is a σ-algebra, it contains $\bigcup_{n \in \mathbb{N}} A_n$, for every α. Therefore, $\Sigma := \bigcap_{\alpha \in J} \Sigma_\alpha$ also contains $\bigcup_{n \in \mathbb{N}} A_n$. ■

Lemma 1.2 shows that $\sigma(X)$ from (1.2) is a σ-algebra.

Example 1.3 (A concrete generated σ-algebra)

Let $\Omega = \{1, 2, 3, 4\}$ and $X = \{\{1, 2\}, \{1, 4\}\}$. We compute $\sigma(X)$ by closing under complements and unions:

$$\{1, 2\},\ \{3, 4\},\ \{1, 4\},\ \{2, 3\},\ \{1, 2, 4\} \to \{1, 2\} \cup \{1, 4\},\ \{3\},\ \{2\},\ \{1\},\ \{4\},\ \ldots \tag{1.3}$$

Continuing this process eventually generates every subset of $\{1, 2, 3, 4\}$, so in this case $\sigma(X) = \mathcal{P}(\Omega)$, with 16 elements.

Example 1.4 (A proper subalgebra)

Let $\Omega = \{1, 2, 3, 4\}$ and $X = \{\{1, 2\}\}$. Closing under complement and union gives

$$\sigma(X) = \{\varnothing,\ \{1, 2\},\ \{3, 4\},\ \Omega\}, \tag{1.4}$$

which has only $4 = 2^2$ elements, which is a proper subset of $\mathcal{P}(\Omega)$ (which has 16 elements). This shows that a σ-algebra need not be the full power set.

Exercise 1.1 (Possible cardinalities)**.** Let $\Omega = \{1, \ldots, n\}$.

(a) Show that every σ-algebra on a finite set Ω has cardinality equal to a power of 2.

(b) For a fixed achievable cardinality 2^k, is the σ-algebra with that cardinality unique? Investigate how many distinct σ-algebras of cardinality 2^k exist on $\{1, \ldots, n\}$.

Solution. Let Σ be a σ-algebra on Ω. We define a relation on Ω as follows:

$$x \sim y \iff \text{for every } A \in \Sigma,\ x \in A \text{ if and only if } y \in A. \tag{1.5}$$

This is clearly an equivalence relation. Hence, the equivalence classes partition Ω. Now, for a given $x \in \Omega$, let C_x be the equivalence class containing x. Then,

$$C_x = \bigcap_{A \in \Sigma,\ x \in A} A \quad \in \Sigma, \tag{1.6}$$

as it's the intersection of sets in Σ. Furthermore, for $A \in \Sigma$,

$$A = \bigcup_{x \in A} C_x. \tag{1.7}$$

Hence, every element of Σ is the disjoint union of members of $\mathcal{C} = \{C_x : x \in \Omega\}$. Furthermore, **every** disjoint union of members of $\mathcal{C}$ is in Σ, since $\mathcal{C} \subseteq \Sigma$. Hence, the elements of Σ are in a bijection with the subsets of $\mathcal{C}$. Therefore,

$$|\Sigma| = 2^{|\mathcal{C}|}. \tag{1.8}$$

For the second part, $|\Sigma| = 2^k$ corresponds to a partition of Ω into k disjoint parts, which is the Stirling number of second kind, denoted as $\left\{ {n \atop k} \right\}$. ■

Example 1.5 (Borel σ-algebra)

Let $\Omega = \mathbb{R}^d$ with its usual topology, and let X be the collection of all **open** subsets of $\mathbb{R}^d$. The generated σ-algebra

$$\mathcal{B}(\mathbb{R}^d) := \sigma(X) \tag{1.9}$$

is called the **Borel σ-algebra**. It contains every open set, every closed set (as complements of open sets), and all sets obtainable from these by countably many unions, intersections, and complements. In particular, on $\mathbb{R}$, it contains every interval (open, closed, or half-open).

It is useful to contrast a σ-algebra with a **topology**, since both are collections of subsets satisfying closure axioms.

Definition 1.2 (Topology). A **topology** τ on a set Ω is a collection of subsets of Ω (called **open sets**) such that:

(i) $\varnothing, \Omega \in \tau$;

(ii) τ is closed under **arbitrary** unions: if $\{O_j\}_{j \in J}$ is any family (not necessarily countable) of sets in τ, then $\bigcup_{j \in J} O_j \in \tau$;

(iii) τ is closed under **finite** intersections: if $O_1, \ldots, O_n \in \tau$, then $\bigcap_{i=1}^{n} O_i \in \tau$.

Remark 1.3 (Key differences).

- A topology allows *arbitrary* (possibly uncountable) unions but only *finite* intersections.
- A σ-algebra allows only *countable* unions (and, as a consequence, countable intersections), but requires closure under **complements**, which a topology does not.

These are genuinely different axiom systems, even though both start from "a set together with a distinguished collection of subsets."

§1.ii. Measures

Having fixed a measurable space (Ω, Σ), we now assign a size to each measurable set.

Definition 1.3 (Measure). A **measure** on (Ω, Σ) is a function $\mu : \Sigma \to \mathbb{R}_{\geq 0} \cup \{\infty\} = [0, \infty]$ such that:

(i) (countable additivity) if $A_1, A_2, A_3, \ldots \in \Sigma$ are pairwise disjoint, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n); \tag{1.10}$$

(ii) μ is not identically $+\infty$ (i.e. $\mu(A) < \infty$ for at least one $A \in \Sigma$).

The triple (Ω, Σ, μ) is called a **measure space**.

Remark 1.4 (Motivation for additivity)**.** If a measure is meant to behave like a physical quantity, e.g. volume of water in a bucket, then combining two *disjoint* quantities should simply add their sizes: 3 liters and 5 liters combined give 8 liters, with nothing created or destroyed. Countable additivity is the natural axiomatization of this idea.

Proposition 1.3 (The measure of the empty set)

Let μ be a measure on (Ω, Σ) satisfying countable additivity (condition (i) above), and suppose μ is not identically $+\infty$. Then $\mu(\varnothing) = 0$.

Proof. Apply additivity to the disjoint sequence $A, \varnothing, \varnothing, \ldots$ for a fixed A with $\mu(A) < \infty$:

$$\mu(A) = \mu(A \cup \varnothing) = \mu(A) + \mu(\varnothing). \tag{1.11}$$

Since $\mu(A) < \infty$, we may subtract it from both sides to get $\mu(\varnothing) = 0$.

(If instead $\mu(\varnothing) = c > 0$ were allowed, one could show $\mu(\varnothing) = \mu(\varnothing \cup \varnothing) = 2c$, forcing $c \in \{0, \infty\}$; and if $\mu(\varnothing) = \infty$ then, applying additivity with A and $\varnothing$, $\mu(A) = \mu(A) + \infty = \infty$ for every A, i.e. μ is identically ∞. This is exactly the degenerate case excluded by (ii).) ■

Remark 1.5. Some textbooks instead simply impose $\mu(\varnothing) = 0$ directly as part of the definition, in place of the "not identically ∞" condition; the two conventions agree on all non-degenerate measures. This mirrors the analogous convention issue for topologies (where $\varnothing, \Omega \in \tau$ is imposed directly).

§1.iii. Finite and σ-finite measures

Definition 1.4. Let (Ω, Σ, μ) be a measure space.

- μ is a **finite measure** if $\mu(\Omega) < \infty$.
- If μ is finite and $\mu(\Omega) = 1$, then μ is called a **probability measure**.
- μ is **σ-finite** if there exists a sequence $A_1, A_2, A_3, \ldots \in \Sigma$ such that

$$\mu(A_n) < \infty \text{ for every } n, \qquad \text{and} \qquad \Omega = \bigcup_{n=1}^{\infty} A_n. \tag{1.12}$$

Remark 1.6. If $\mu(\Omega) < \infty$, one can always **normalize** μ by setting $\tilde{\mu}(A) = \mu(A)/\mu(\Omega)$ (assuming $\mu(\Omega) > 0$), producing a probability measure with the same measurable sets of measure zero.

Proposition 1.4 (Continuity of measure)

Let (M, Σ, μ) be a measure space.

(a) If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of measurable sets, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu\left(A_n\right). \tag{1.13}$$

(b) If $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of measurable sets and $\mu\left(A_1\right) < \infty$, then

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu\left(A_n\right). \tag{1.14}$$

Proof. (a) We define $B_1 = A_1$ and $B_n = A_n - A_{n-1}$ for $n \geq 2$. Then the B_n's are pairwise disjoint, and $\bigcup_n B_n = \bigcup_n A_n$. As a result,

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu\left(B_n\right) \tag{1.15}$$

$$= \lim_{N\to\infty} \sum_{n=1}^{N} \mu\left(B_n\right) = \lim_{N\to\infty} \mu\left(\bigcup_{n=1}^{N} B_n\right) \tag{1.16}$$

$$= \lim_{N\to\infty} \mu\left(A_N\right), \tag{1.17}$$

since $\bigcup_{n=1}^{N} B_n = A_N$.

(b) Define $B_n = A_1 - A_n$. Then B_n is an increasing sequence of measurable sets. Furthermore,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} (A_1 - A_n) = A_1 - \bigcap_{n=1}^{\infty} A_n. \tag{1.18}$$

Claim 1.1: If A, B are measurable sets with $A \subseteq B$ and $\mu(A) < \infty$, then $\mu\left(B - A\right) = \mu(B) - \mu(A)$.

Proof. $B = A \cup (B - A)$ and this is a disjoint decomposition. Therefore,

$$\mu(B) = \mu(A) + \mu\left(B - A\right). \tag{1.19}$$

Rearranging gives the desired result. □

By applying part (a), we get

$$\mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu\left(B_n\right). \tag{1.20}$$

Applying Claim 1.1 into both sides of the equation, we get

$$\mu(A_1) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n). \tag{1.21}$$

■

Remark 1.7. The condition that $\mu(A_1) < \infty$ (or at least $\mu(A_n) < \infty$ for some n) is important here. For example, consider

$$\Omega = \mathbb{N}, \quad \Sigma = \mathcal{P}(\mathbb{N}), \quad \mu = \text{ counting measure }.$$

Here, $A_n = \{n, n+1, n+2, \ldots\}$ is a decreasing sequence with $\mu(A_n) = \infty$ for every n. But

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \mu(\varnothing) = 0 \text{ and } \lim_{n\to\infty} \mu(A_n) = \infty. \tag{1.22}$$

Example 1.6

Lebesgue measure on $\mathbb{R}^d$ is not finite (the whole space has infinite volume), but it *is* σ-finite: taking $A_n = B(0, n)$, the ball of radius n centered at the origin, each A_n has finite volume and $\mathbb{R}^d = \bigcup_{n=1}^{\infty} A_n$.

Remark 1.8. The choice of Ω, Σ, and μ is entirely up to the application at hand. There is no canonical choice in general. However, certain combinations are used so pervasively that they become the default setting for the theory; the main example developed in this course will be $\mathbb{R}^d$ equipped with the Borel (or Lebesgue) σ-algebra and Lebesgue measure.

§1.iv. Measurable Functions

Definition 1.5 (Measurable function)**.** Let (Ω, Σ) be a measurable space. A function $f : \Omega \longrightarrow \mathbb{R}$ is called **measurable** (with respect to Σ) if, for every interval $I \subseteq \mathbb{R}$,

$$f^{-1}(I) := \{x \in \Omega : f(x) \in I\} \in \Sigma. \tag{1.23}$$

Remark 1.9. It suffices to check this condition for intervals of the special form $I = (a, \infty)$, $a \in \mathbb{R}$: since Σ is closed under complements, countable unions, and countable intersections, every interval (open, closed, half-open, bounded, or unbounded) can be built from intervals of this one

type. Because,

$$(-\infty, a] = (a, \infty)^c\,, \quad [a, \infty) = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, \infty\right), \quad (-\infty, a) = [a, \infty)^c\,,$$
$$(a, b) = (a, \infty) \cap (-\infty, b)\,, \quad [a, b] = [a, \infty) \cap (-\infty, b]\,,$$
$$[a, b) = [a, \infty) \cap (-\infty, b)\,, \quad (a, b] = (a, \infty) \cap (-\infty, b]\,;$$

and preimage distributes union, intersections and complements. That's why many texts simply *define* measurability by requiring $f^{-1}((a, \infty)) \in \Sigma$ for all $a \in \mathbb{R}$.

Remark 1.10 (Measurability depends on the choice of Σ)**.** Whether a given function is measurable depends entirely on the σ-algebra chosen on the domain. It is not an intrinsic property of the function alone.

Example 1.7

Let $\Omega = \mathbb{R}$ and $\Sigma = \{\varnothing, \mathbb{R}\}$ (the trivial σ-algebra).

- The constant function $f(x) = 1$ is measurable: for any interval I, $f^{-1}(I)$ is either $\mathbb{R}$ (if $1 \in I$) or $\varnothing$ (if $1 \notin I$), and both lie in Σ.
- The identity function $f(x) = x$ is **not** measurable: for a bounded interval such as $I = (0, 1)$, $f^{-1}(I) = (0, 1) \notin \{\varnothing, \mathbb{R}\}$.

This illustrates that measurability is relative to Σ: the same function can be measurable for one σ-algebra and not for another.

Example 1.8 (Basic examples of measurable functions)

The following are basic examples of measurable functions (with respect to the Borel σ-algebra, unless otherwise specified):

(i) Any constant function $f \equiv c$ is measurabl.

(ii) For a measurable set $A \subseteq \Omega$, the **characteristic function** of A is defined by

$$\chi_A(x) := \begin{cases} 1, & \text{if } x \in A; \\ 0, & \text{if } x \notin A. \end{cases} \tag{1.24}$$

This function is measurable: for $a \in \mathbb{R}$, the following sets are all measurable

$$\{x \in \Omega : \chi_A(x) > a\} = \begin{cases} \Omega, & a < 0; \\ A, & 0 \le a < 1; \\ \varnothing, & a \ge 1. \end{cases} \tag{1.25}$$

(iii) Every continuous function $f : \mathbb{R}^d \to \mathbb{R}$ is measurable with respect to the Borel σ-

algebra: for any $a \in \mathbb{R}$, the set $\{x : f(x) > a\} = f^{-1}((a, \infty))$ is open (being the preimage of an open set under a continuous map), hence Borel measurable.

(iv) Every monotone function $f : \mathbb{R} \to \mathbb{R}$ (increasing or decreasing) is measurable with respect to the Borel σ-algebra: for any $a \in \mathbb{R}$, the set $\{x : f(x) > a\}$ is always an interval (a ray, of one of the forms (c, ∞), $[c, \infty)$, $(-\infty, c)$, or $(-\infty, c]$, depending on monotonicity and endpoint conventions), and every interval is Borel measurable.

Remark 1.11 (Measurable maps between measure spaces)**.** More generally, if (Ω_1, Σ_1) and (Ω_2, Σ_2) are two measurable spaces, a function $f : \Omega_1 \to \Omega_2$ is called **measurable** if $f^{-1}(B) \in \Sigma_1$ for every $B \in \Sigma_2$. The definition above is the special case $\Omega_2 = \mathbb{R}$ with $\Sigma_2 = \mathcal{B}(\mathbb{R})$, the Borel σ-algebra.

It will be convenient for us to also allow our functions to take values in the extended real line $\overline{\mathbb{R}}$. The definition is the same as before: a function $f : \Omega \to \overline{\mathbb{R}}$ is **measurable** if the preimage of any extended interval is measurable, or, alternatively, if the sets $\{x \in \Omega : f(x) > a\}$ are measurable for each $a \in \mathbb{R}$. In this case, observe that the sets

$$\{x \in \Omega : f(x) = +\infty\} = \bigcap_{n=1}^{\infty} \{x \in \Omega : f(x) > n\}, \tag{1.26}$$

$$\{x \in \Omega : f(x) = -\infty\} = \left(\bigcup_{n=1}^{\infty} \{x \in \Omega : f(x) > -n\} \right)^c \tag{1.27}$$

are both measurable.

We shall adopt the convention that

$$0 \cdot (\pm\infty) = 0 \text{ and } (\pm infty) + \pm (\mp\infty) = 0. \tag{1.28}$$

Let $\mathcal{M} = \mathcal{M}(\Omega, \Sigma)$ denote the set of measurable functions $f : \Omega \to \mathbb{R}$ (or $\overline{\mathbb{R}}$).

Proposition 1.5

Let $f, g \in \mathcal{M}$ and $c \in \mathbb{R}$. Then each of the following functions is also measurable:

(a) cf;

(b) $f \pm g$;

(c) fg;

(d) $\frac{f}{g}$, if $g \neq 0$ everywhere;

(e) $|f|$.

Proof. We are going to use the trick that f is measurable if the sets $\{x \in \Omega : f(x) > a\}$ are all measurable.

(a) If $c = 0$, then $cf = 0$, which is a constant function, so it's measurable. For $c > 0$,

$$\{x \in \Omega : c \cdot f(x) > a\} = \left\{x \in \Omega : \cdot f(x) > \frac{a}{c}\right\} \tag{1.29}$$

is measurable since f is a measurable function. For $c < 0$,

$$\{x \in \Omega : c \cdot f(x) > a\} = \left\{x \in \Omega : \cdot f(x) < \frac{a}{c}\right\} \tag{1.30}$$

is measurable since f is a measurable function.

(b) Note that (assuming $f(x), g(x)$ are finite)

$$\begin{aligned} f(x) + g(x) > a &\iff f(x) > a - g(x) \\ &\iff \exists q \in \mathbb{Q} \text{ such that } f(x) > q > a - g(x) \\ &\iff \exists q \in \mathbb{Q} \text{ such that } f(x) > q \text{ and } g(x) > a - q. \end{aligned}$$

$$\begin{aligned} &\{x \in \Omega : f(x) + g(x) > a\} \\ = &\{x \in \Omega : \text{ both } f(x) \text{ and } g(x) \text{ are finite, and } f(x) + g(x) > a\} \\ &\cup (\{x \in \Omega : f(x) = \infty\} \cap \{x \in \Omega : g(x) > -\infty\}) \\ &\cup (\{x \in \Omega : g(x) = \infty\} \cap \{x \in \Omega : f(x) > -\infty\}) \\ &\cup X, \end{aligned} \tag{1.31}$$

where X is

$$X = \begin{cases} \{x \in \Omega : f(x) = +\infty, g(x) = -\infty\} \cup \{x \in \Omega : g(x) = +\infty, f(x) = -\infty\}, & a < 0, \\ \varnothing, & a \geq 0. \end{cases} \tag{1.32}$$

Therefore,

$$\begin{aligned} &\{x \in \Omega : f(x) + g(x) > a\} \\ = &\bigcup_{q \in \mathbb{Q}} (\{x \in \Omega : f(x) > q\} \cap \{x \in \Omega : g(x) > a - q\} \cap \{x \in \Omega : f(x), g(x) \in \mathbb{R}\}) \\ &\cup (\{x \in \Omega : f(x) = \infty\} \cap \{x \in \Omega : g(x) > -\infty\}) \\ &\cup (\{x \in \Omega : g(x) = \infty\} \cap \{x \in \Omega : f(x) > -\infty\}) \\ &\cup X. \end{aligned} \tag{1.33}$$

This is a measurable set since it's made up of measurable sets by taking union and intersection.

(c) We're gonna prove an intermediate lemma first.

> **Claim 1.2:** If f is measurable, then so is f^2.

Proof. If $a < 0$, then $\left\{x \in \Omega : f(x)^2 > a\right\} = \Omega$, is measurable. For $a \geq 0$,

$$\left\{x \in \Omega : f(x)^2 > a\right\} = \left\{x \in \Omega : f(x) > \sqrt{a}\right\} \cup \left\{x \in \Omega : f(x) < -\sqrt{a}\right\} \tag{1.34}$$

is measurable. □

Now notice that

$$fg = \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right], \tag{1.35}$$

is indeed measurable.

(d) It suffices to show that $\frac{1}{g}$ is measurable.

$$\left\{x \in \Omega : \frac{1}{g(x)} > a\right\} = \begin{cases} \left\{x \in \Omega : 0 < g(x) > \frac{1}{a}\right\}, & a > 0, \\ \left\{x \in \Omega : g(x) > 0\right\}, & a = 0, \\ \left\{x \in \Omega : \frac{1}{a} < g(x) < 0\right\}, & a < 0. \end{cases} \tag{1.36}$$

These are all measurable since g is measurable.

(e) If $a < 0$, then $\{x \in \Omega : |f(x)| > a\} = \Omega$, is measurable. For $a \geq 0$,

$$\{x \in \Omega : |f(x)| > a\} = \{x \in \Omega : f(x) > a\} \cup \{x \in \Omega : f(x) < -a\} \tag{1.37}$$

is measurable.

■

§1.v. Basic properties of measurable functions

Proposition 1.6

Let $(f_n)_{n=1}^{\infty}$ be a countable sequence of measurable functions $\Omega \to \overline{\mathbb{R}}$. Then each of the following functions is measurable:

(a) $f(x) = \sup_n f_n(x)$;

(b) $f(x) = \inf_n f_n(x)$;

(c) $f(x) = \limsup_{n\to\infty} f_n(x)$;

(d) $f(x) = \liminf_{n\to\infty} f_n(x)$.

Consequently, if $f_n \to f$ pointwise, then f is measurable.

Proof. (a) Note that

$$\sup_n f_n(x) > a \iff f_n(x) > a \text{ for some } n. \tag{1.38}$$

Therefore,

$$\left\{x \in \Omega : \sup_n f_n(x) > a\right\} = \bigcup_n \{x \in \Omega : f_n(x) > a\}, \tag{1.39}$$

is measurable.

(b) Note that

$$\inf_n f_n(x) = -\left(\sup_n \left(-f_n(x)\right)\right). \tag{1.40}$$

(c) Follows from the fact that

$$\limsup_{n\to\infty} f_n(x) = \inf_m \sup_{n\geq m} f_n(x). \tag{1.41}$$

(d) Follows from the fact that

$$\liminf_{n\to\infty} f_n(x) = \sup_m \inf_{n\geq m} f_n(x). \tag{1.42}$$

Whem $f_n \to f$ pointwise, then the limit, limsup and liminf coincides. ■

Remark 1.12. In the case of complex-valued functions, we say that $f : \Omega \to \mathbb{C}$ is measurable if its real part and imaginary parts are both measurable.

Definition 1.6 (Almost everywhere)**.** Let (Ω, Σ, μ) be a measure space. We say that a property (P) holds μ-**almost everywhere** (or simply **almost everywhere**, when the measure μ is clear from context), abbreviated μ-a.e. or a.e., if there exists a set $N \in \Sigma$ with $\mu(N) = 0$ such that (P) holds on N^c. Such a set N is called a **null set**.

Example 1.9

Typical almost-everywhere statements include:

- $f = g$ almost everywhere, meaning $\{x \in \Omega : f(x) \neq g(x)\}$ is contained in a null set;
- $\lim_{n\to\infty} f_n(x) = f(x)$ almost everywhere, meaning the set of x for which this limit fails to hold is contained in a null set.

Remark 1.13 (Subsets of null sets need not be measurable)**.** If $\mu(N) = 0$ and $B \subseteq N$, one would like to conclude that $\mu(B) = 0$ as well. However, this conclusion presupposes $B \in \Sigma$, which need not be the case, since Σ may fail to contain every subset of a null set.

§1.vi. Complete measure

Definition 1.7 (Complete measure)**.** A measure μ on (Ω, Σ) is called **complete** if its domain Σ contains every subset of every μ-null set; that is, whenever $N \in \Sigma$ satisfies $\mu(N) = 0$ and

$B \subseteq N$, then $B \in \Sigma$ (and consequently $\mu(B) = 0$).

Remark 1.14. Every measure can be made complete by slightly enlarging its σ-algebra, as follows. Let (Ω, Σ, μ) be a measure space. Let

$$\mathcal{N} := \{N \in \Sigma : \mu(N) = 0\}, \qquad \overline{\Sigma} := \{A \cup B : A \in \Sigma \text{ and } B \subseteq N \text{ for some } N \in \mathcal{N}\}. \tag{1.43}$$

$\overline{\Sigma}$ is a σ-algebra. Indeed:

1. Take $A \cup B \in \overline{\Sigma}$, where $A \in \Sigma$ and $B \subseteq N$ for some $N \in \mathcal{N}$. Then

$$(A \cup B)^c = A^c \cap B^c = A^c - (B \cap A^c)\,. \tag{1.44}$$

   We're gonna use $W - U = (W - V) \cup (V - U)$ for $U \subseteq V \subseteq W$, with $U = B \cap A^c$, $V = N \cap A^c$, $W = A^c$. Then the RHS of (1.44) becomes,

$$[A^c - (N \cap A^c)] \cup [(N \cap A^c) - (B \cap A^c)]\,. \tag{1.45}$$

   $A^c - (N \cap A^c) \in \Sigma$ since it's made up of elements of Σ using complementation and finite intersection. $(N \cap A^c) - (B \cap A^c) \subseteq N$. Therefore, $(A \cup B)^c \in \overline{\Sigma}$.

2. Let $A_i \cup B_i \in \overline{\Sigma}$ be a countable sequence of sets, with $A_i \in \Sigma$ and $B_i \subseteq N_i$ with $\mu(N_i) = 0$. Then

$$\mu\left(\bigcup_{i=1}^{\infty} N_i\right) \leq \sum_{i=1}^{\infty} \mu(N_i) = 0, \tag{1.46}$$

   implying that $\bigcup_{i=1}^{\infty} N_i \in \mathcal{N}$. Now,

$$\bigcup_{i=1}^{\infty} (A_i \cup B_i) = \left(\bigcup_{i=1}^{\infty} A_i\right) \cup \left(\bigcup_{i=1}^{\infty} B_i\right), \tag{1.47}$$

   where $\bigcup_{i=1}^{\infty} A_i \Sigma$, and $\bigcup_{i=1}^{\infty} B_i \subseteq \bigcup_{i=1}^{\infty} N_i \in \mathcal{N}$.

There is a unique extension $\overline{\mu}$ of μ to a complete measure on $\overline{\Sigma}$, given by

$$\overline{\mu}(A \cup B) = \mu(A)\,, \tag{1.48}$$

where $A \in \Sigma$ and $B \subseteq N$ for some $N \in \mathcal{N}$. The measure space $(\Omega, \overline{\Sigma}, \overline{\mu})$ is called the **completion** of (Ω, Σ, μ).

Proposition 1.7

$\overline{\mu}$ is a complete measure, and it's the unique measure on $\overline{\Sigma}$ that extends μ.

Proof. Suppose $\mu(A \cup B) = 0$ for some $A \cup B \in \overline{\Sigma}$, where $A \in \Sigma$ and $B \subseteq N$ with $\mu(N) = 0$. Choose any subset $C \subseteq A \cup B$. We can write it as

$$C = \varnothing \cup C, \text{ for } C \subseteq A \cup N, \tag{1.49}$$

where $A \cup N \in \Sigma$ with $\mu(A \cup N) = 0$. Therefore, $C \in \overline{\Sigma}$, and hence $\overline{\mu}$ is a complete measure.

Assume for the sake of contradiction that μ' is a measure on $\overline{\Sigma}$ that extends μ, i.e. $\mu'|_{\Sigma} = \mu$. For $A \cup B \in \overline{\Sigma}$, where $A \in \Sigma$ and $B \subseteq N$ with $\mu(N) = 0$, we have

$$\overline{\mu}(A \cup B) = \mu(A) = \mu'(A) \leq \mu'(A \cup B) \leq \mu'(A \cup N) = \mu(A \cup N) = \mu(A) = \overline{\mu}(A \cup B). \tag{1.50}$$

Therefore, $\mu' = \overline{\mu}$. ■

Lemma 1.8

The following implications hold **if and only if** the measure μ is complete:

(i) if f is measurable and $f = g$ μ-a.e., then g is measurable;

(ii) if $(f_n)_{n \geq 1}$ are measurable and $f_n \to f$ μ-a.e., then f is measurable.

Proof. (i) Suppose μ is complete; and f is measurable and $f = g$ a.e. Then $f \neq g$ on $B \subseteq N$, where $\mu(N) = 0$. Since μ is complete, $B \in \Sigma$. Now,

$$\{x \in \Omega : g(x) > a\} = (\{x \in \Omega : g(x) > a\} \cap B) \cup (\{x \in \Omega : g(x) > a\} \cap B^c) \tag{1.51}$$

$$= (\text{subset of a null-set}) \cup (\{x \in \Omega : f(x) > a\} \cap B^c), \tag{1.52}$$

which is a measurable set.

Conversely, suppose that whenever f is measurable and $f = g$ a.e., we have that g is measurable. Now, let $N \in \mathcal{N}$ and $B \subseteq N$. Take

$$f \equiv 0, \quad g = \chi_B \tag{1.53}$$

They agree on N^c, so $f = g$ a.e. f is a constant function so it's measurable. Therefore, g is also measurable, which forces

$$g^{-1}\{1\} = B \in \Sigma. \tag{1.54}$$

(ii) Suppose $f_n \to g$ (everywhere). Then g is measurable, and $f = g$ a.e. Therefore, f is measurable if and only if μ is complete.

■

Lemma 1.9 (Borel–Cantelli lemma)

Let (Ω, Σ, μ) be a measure space. Let $(E_k)_{k=1}^{\infty}$ be a countable family of measurable sets of Ω such that

$$\sum_{k=1}^{\infty} \mu(E_k) < \infty. \tag{1.55}$$

Let $E := \{x \in \Omega : x \in E_k \text{ for infinitely many } k\}$. Then E is measurable and $\mu(E) = 0$.

Proof. Note that

$$\begin{aligned} x \in E_k \text{ for infinitely many } k &\iff \forall n\ \exists m \geq n \text{ such that } x \in E_m \\ &\iff \forall n, x \in \bigcup_{m \geq n} E_m \\ &\iff x \in \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} E_m. \end{aligned}$$

Therefore, $E = \bigcap_{n=1}^{\infty} \bigcup_{m \geq n} E_m$ is measurable as it's made up of measurable sets by counrable unions and intersections. Now,

$$\mu\left(\bigcup_{m \geq 1} E_m\right) \leq \sum_{m=1}^{\infty} \mu\left(E_m\right) < \infty. \tag{1.56}$$

By continuity, since $\left\{\bigcup_{m \geq n} E_m\right\}_n$ is a decreasing sequence,

$$\mu\left(E\right) = \lim_{n \to \infty} \mu\left(\bigcup_{m \geq n} E_m\right) \leq \lim_{n \to \infty} \sum_{m=n}^{\infty} \mu\left(E_m\right) = 0. \tag{1.57}$$

■

§1.vii. Exercises

Exercise 1.2. Let $(f_n)_{n \geq 1}$ be a sequence of (measurable) functions on a measure space (Ω, Σ, μ). Suppose that for every $x \in \Omega$, the sequence $(f_n(x))_{n \geq 1}$ is bounded. Let

$$E := \{x \in \Omega : (f_n(x))_{n \geq 1} \text{ converges}\}. \tag{1.58}$$

Show that $E \in \Sigma$.

Solution. $(f_n(x))_{n \geq 1}$ converges if and only if $\limsup_n f_n(x) = \liminf_n f_n(x)$. We have already shown that $\limsup_n f_n$ and $\liminf_n f_n$ are measurable. Therefore, so is $h := \limsup_n f_n - \liminf_n f_n$. Now,

$$E = \{x \in \Omega : (f_n(x))_{n \geq 1} \text{ converges}\} = h^{-1}\left\{0\right\}, \tag{1.59}$$

which is measurable since h is measurable. ■

Exercise 1.3. For each $n \geq 1$, let $f_n : [0,1] \to \mathbb{R}$ be a continuous function. Suppose that:

(i) $0 \leq f_n(x)$ for all n and all $x \in [0,1]$;

(ii) the sequence of partial sums $\sum_{n=1}^{N} f_n(x)$ converges uniformly on $[0,1]$ as $N \to \infty$;

(iii) the sum $\displaystyle\sum_{n=1}^{\infty} \left(\sup\{f_n(x) : 0 \leq x \leq 1\}\right)$ diverges.

Either give an example of such a sequence of functions, or prove that no such sequence exists.

Solution. One such sequence of functions is

$$f_n(x) = \text{ triangular bump function on } \left[\frac{1}{n+1}, \frac{1}{n}\right] \text{ of height } \frac{1}{n}. \tag{1.60}$$

Then

$$\sum_{n=1}^{\infty} (\sup\{f_n(x) : 0 \le x \le 1\}) = \sum_{n=1}^{\infty} \frac{1}{n} \text{ diverges .} \tag{1.61}$$

$$\sum_{n=1}^{N} f_n(x) = \text{ triangular bump functions of height } \frac{1}{n} \text{ on each } \left[\frac{1}{n+1}, \frac{1}{n}\right], \text{ for } n = 1, 2, \ldots, N. \tag{1.62}$$

which converges uniformly to

$$f(x) = \begin{cases} \text{triangular bump functions of height } \frac{1}{n} & \text{if } x \in \left[\frac{1}{n+1}, \frac{1}{n}\right] \text{ for } n \ge 1, \\ 0 & x = 0. \end{cases} \tag{1.63}$$

Indeed,

$$\sup_{x\in[0,1]} \left| f(x) - \sum_{n=1}^{N} f_n(x) \right| = \frac{1}{N+1} \to 0. \tag{1.64}$$

■

§2. 11 September 2026

§2.i. Motivation and Characteristic Functions

Having built the machinery of σ-algebras, measures, and measurable functions, we now turn to the central question of this part of the course:

Given a measure space (Ω, Σ, μ) and a measurable function f, how do we define $\displaystyle\int_\Omega f \, \mathrm{d}\mu$?

The construction proceeds in three stages: first for characteristic functions, then for simple functions, and finally for general non-negative measurable functions (by approximation). Integration of signed or complex-valued functions is deferred to a later lecture.

Definition 2.1 (Characteristic function)**.** Let $A \in \Sigma$ be a measurable set. The **characteristic function** (or **indicator function**) of A is the function $\chi_A : \Omega \to \mathbb{R}$ (also written $\mathbf{1}_A$) defined by

$$\chi_A(x) := \begin{cases} 1, & x \in A; \\ 0, & x \notin A. \end{cases} \tag{2.1}$$

Remark 2.1. As noted already, χ_A is always measurable when $A \in \Sigma$: for an interval $I \subseteq \mathbb{R}$,

$$\chi_A^{-1}(I) = \begin{cases} \Omega, & 0, 1 \in I; \\ A, & 1 \in I,\ 0 \notin I; \\ A^c, & 0 \in I,\ 1 \notin I; \\ \varnothing, & 0, 1 \notin I, \end{cases} \tag{2.2}$$

and each of $\Omega, A, A^c, \varnothing$ lies in Σ.

The entire theory of integration is built by insisting on one basic axiom, motivated by the idea that the integral should compute the "volume" below the graph of a function: for a measurable set A and a constant height $c \geq 0$, the volume of the block of height c over the base A should be c times the size of the base. This gives the defining relation

$$\int_\Omega c\,\chi_A \, \mathrm{d}\mu := c\,\mu(A), \qquad c \geq 0,\ A \in \Sigma, \tag{2.3}$$

and, in particular (taking $c = 1$),

$$\int_\Omega \chi_A \, \mathrm{d}\mu := \mu(A). \tag{2.4}$$

§2.ii. Simple Functions

Definition 2.2 (Simple function)**.** A measurable function $\varphi : \Omega \to \mathbb{R}$ is called **simple** if it takes only finitely many values.

Remark 2.2 (Standard representation)**.** If φ is simple and $a_1, \ldots, a_N$ are its distinct *non-zero* values, set

$$A_i := \varphi^{-1}(\{a_i\}), \qquad i = 1, \ldots, N. \tag{2.5}$$

Each A_i is measurable (since φ is measurable), the sets $A_1, \ldots, A_N$ are pairwise disjoint, and

$$\varphi = \sum_{i=1}^{N} a_i \, \chi_{A_i}. \tag{2.6}$$

This expression is called the **standard representation** of φ.

Remark 2.3 (Non-standard representations)**.** A simple function can typically be written as a finite linear combination of characteristic functions in more than one way, using sets that are *not* disjoint and coefficients that are *not* distinct. For example, on disjoint sets A, B, C with

$$\varphi = 5\,\chi_A + 8\,\chi_B + 3\,\chi_C \tag{2.7}$$

(the standard representation), one may check directly that

$$\varphi = 5\,\chi_{A\cup B} + 3\,\chi_{B\cup C} \tag{2.8}$$

as well, since on A this gives 5, on B it gives $5 + 3 = 8$, and on C it gives 3 — matching the standard representation, even though $A \cup B$ and $B \cup C$ overlap.

Definition 2.3 (Integral of a non-negative simple function)**.** Let φ be a non-negative simple function with standard representation $\varphi = \sum_{i=1}^{N} a_i \chi_{A_i}$ (so $a_i > 0$ and the A_i are pairwise disjoint, possibly with $\mu(A_i) = \infty$). We define

$$\int_\Omega \varphi \, \mathrm{d}\mu := \sum_{i=1}^{N} a_i \, \mu(A_i) \ \in [0, \infty]. \tag{2.9}$$

Lemma 2.1 (Independence of the representation)

Let $\varphi \geq 0$ be a simple function, and suppose

$$\varphi = \sum_{n=1}^{N} a_n \chi_{A_n} = \sum_{m=1}^{M} b_m \chi_{B_m} \tag{2.10}$$

are two representations of φ as finite linear combinations of characteristic functions, where the sets A_n (resp. B_m) need *not* be pairwise disjoint, and the coefficients $a_n, b_m \in \mathbb{R}$ need *not* be distinct or non-negative (only φ itself is assumed non-negative; we adopt the convention

$0 \times \infty = 0$ if some $\mu(A_n) = \infty$ with $a_n = 0$). Then

$$\sum_{n=1}^{N} a_n \, \mu(A_n) = \sum_{m=1}^{M} b_m \, \mu(B_m). \tag{2.11}$$

Proof. We'll show that $\sum_{n=1}^{N} a_n \, \mu(A_n) = \sum_{i=1}^{K} r_i \, \mu(E_i)$, where $\varphi = \sum_{i=1}^{K} r_i \, \chi_{E_i}$ is the standard representation.

Case 1: A_i's are pairwise disjoint, but a_i's need not be distinct. In this case, for each $a \in \{a_1, \ldots, a_N\}$, set $B_a = \bigcup_{a_i = a} A_i$. Then it's clear that $B_a \cap B_{a'} = \varnothing$, since A_i's are pairwise disjoint. Furthermore, $\varphi^{-1}(a) = B_a$. Therefore, $\varphi = \sum_a a \chi_{B_a}$ is the standard representation.

$$\sum_a a\mu(B_a) = \sum_a a \sum_{a_i = a} \mu(A_i) = \sum_{i=1}^{N} a_i \mu(A_i). \tag{2.12}$$

Case 2: Now let's assume that A_i's are not necessarily pairwise disjoint.

Claim 2.1: There exists a refinement of the set $\bigcup_{i=1}^{N} A_i$ with sets $B_1, \ldots, B_M$ such that $\bigcup_{i=1}^{N} A_i = \bigcup_{j=1}^{M} B_j$, and B_j's are pairwise disjoint, and for each i, $A_i = \bigcup_{B_j \subseteq A_i} B_j$.

Proof. For each nonempty subset S of $\{1, \ldots, N\}$, we set

$$B_S = \left(\bigcap_{i \in S} A_i \right) \cap \left(\bigcap_{i \notin S} A_i^c \right). \tag{2.13}$$

This gives a collection of $2^N - 1$ pairwise disjoint sets. Now, we'll show that

$$A_i = \bigcup_{B_S \subseteq A_i} B_S. \tag{2.14}$$

Clearly, $\bigcup_{B_S \subseteq A_i} B_S \subseteq A_i$ holds. Now, take $x \in A_i$. Let $S' = \{i : x \in A_i\}$, which is a nonempty set. Then $x \in B_{S'} \subseteq A_i$ since S' contains i. Therefore, $\bigcup_{B_S \subseteq A_i} B_S \supseteq A_i$.

This proves $\bigcup_{S \neq \varnothing} B_S \supseteq \bigcup_{i=1}^{N} A_i$. The reverse containment holds since $B_S \subseteq \bigcup_{i=1}^{N} A_i$. □

Now, for each S, we define $b_S = \sum_{A_i \supseteq B_S} a_i$. Then

$$\sum_{i=1}^{N} a_i \mu(A_i) = \sum_{i=1}^{N} a_i \sum_{B_S \subseteq A_i} \mu(B_S) = \sum_S \sum_{B_S \subseteq A_i} a_i \, \mu(B_S) = \sum_S b_S \mu(B_S). \tag{2.15}$$

This boils down to case 1, and we know that this is equal to the sum in the standard representation.

■

Remark 2.4. The integral of a non-negative simple function may be computed using *any* representation of φ as a finite linear combination of characteristic functions of measurable sets,

not only the standard one. Equivalently, a non-negative simple function may be described as any finite non-negative linear combination of characteristic functions of measurable sets, and its integral is unambiguously

$$\int_\Omega \varphi \, \mathrm{d}\mu = \sum_n a_n \, \mu(A_n) \tag{2.16}$$

for any such representation $\varphi = \sum_n a_n \chi_{A_n}$.

Lemma 2.2

Let φ and ψ be simple functions in $\mathcal{M}^+$.

(a) If $c \in \mathbb{R}$, then $\int c\varphi \, \mathrm{d}\mu = c \int \varphi \, \mathrm{d}\mu$.

(b) $\int (\varphi + \psi) \, \mathrm{d}\mu = \int \varphi \, \mathrm{d}\mu + \int \psi \, \mathrm{d}\mu$.

(c) If $\varphi \leq \psi$, then $\int \varphi \, \mathrm{d}\mu \leq \int \psi \, \mathrm{d}\mu$.

(d) $A \mapsto \int_A \varphi \, \mathrm{d}\mu := \int_\Omega \varphi \cdot \chi_A \, \mathrm{d}\mu$ is a measure on Σ.

Proof. (a) Let $\varphi = \sum_{i=1}^n a_i \chi_{A_i}$. Then $c\varphi = \sum_{i=1}^n c a_i \chi_{A_i}$.

$$\int_\Omega c\varphi \, \mathrm{d}\mu = \sum_{i=1}^n c a_i \mu\left(A_i\right) = c \sum_{i=1}^n a_i \mu\left(A_i\right) = c \int_\Omega \varphi \, \mathrm{d}\mu. \tag{2.17}$$

(b) Let $\varphi = \sum_{i=1}^n a_i \chi_{A_i}$ and $\psi = \sum_{j=1}^m b_j \chi_{B_j}$. Then $\varphi + \psi = \sum_{i=1}^n a_i \chi_{A_i} + \sum_{j=1}^m b_j \chi_{B_j}$.

$$\int_\Omega (\varphi + \psi) \, \mathrm{d}\mu = \sum_{i=1}^n a_i \mu(A_i) + \sum_{j=1}^m b_j \mu(B_j) = \int_\Omega \varphi \, \mathrm{d}\mu + \int_\Omega \psi \, \mathrm{d}\mu. \tag{2.18}$$

(c) $\psi - \varphi$ is a non-negative simple function, whose integral is non-negative, by definition.

(d) Let $\varphi = \sum_{i=1}^n a_i \chi_{E_i}$ be the standard representation.

$$\nu\left(A\right) := \int_\Omega \varphi \cdot \chi_A \, \mathrm{d}\mu = \int_\Omega \sum_{i=1}^n a_i \chi_{E_i} \cdot \chi_A \, \mathrm{d}\mu = \int_\Omega \sum_{i=1}^n a_i \chi_{E_i \cap A} \, \mathrm{d}\mu = \sum_{i=1}^n a_i \mu\left(E_i \cap A\right). \tag{2.19}$$

Clearly, $\nu\left(A\right) \geq 0$ and $\nu\left(\varnothing\right) = 0$. Now, if $\{A_n\}_n$ is a countable collection of pairwise disjoint sets,

$$\nu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{i=1}^n a_i \mu\left(E_i \cap \bigcup_{n=1}^\infty A_n\right) = \sum_{i=1}^n a_i \mu\left(\bigcup_{n=1}^\infty \left(E_i \cap A_n\right)\right) \tag{2.20}$$

$E_i \cap A_n$'s are all pairwise disjoint, so

$$\nu\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{i=1}^n a_i \sum_{n=1}^\infty \mu\left(E_i \cap A_n\right) = \sum_{n=1}^\infty \sum_{i=1}^n a_i \mu\left(E_i \cap A_n\right) = \sum_{n=1}^\infty \nu\left(E_i \cap A_n\right). \tag{2.21}$$

■

§2.iii. Approximation of Non-negative Functions by Simple Functions

We now turn to general non-negative measurable functions. The key technical tool, used repeatedly throughout analysis, is to approximate an object in a large class (here, non-negative measurable functions) by objects in a smaller class we already understand (here, simple functions).

Theorem 2.3 (Approximation by simple functions)

Let $f : \Omega \to [0, \infty]$ be measurable. Then there exists a sequence of simple functions $(\varphi_n)_{n\geq 1}$, each satisfying $0 \leq \varphi_n \leq \varphi_{n+1}$, such that

$$\varphi_n(x) \longrightarrow f(x) \qquad \text{for every } x \in \Omega. \tag{2.22}$$

Proof. For each $n \geq 1$ and each integer k with $0 \leq k \leq n2^n - 1$, define the measurable set

$$A_{k,n} := \left\{ x \in \Omega : \frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n} \right\}, \tag{2.23}$$

and, to handle the top of the range, set

$$A_{n2^n,n} := \{x \in \Omega : f(x) \geq n\}. \tag{2.24}$$

These sets are measurable (being preimages of intervals under f) and, for each fixed n, pairwise disjoint, with

$$\Omega = \bigcup_{k=0}^{n2^n} A_{k,n}. \tag{2.25}$$

Define

$$\varphi_n := \sum_{k=0}^{n2^n} \frac{k}{2^n}\, \chi_{A_{k,n}}, \tag{2.26}$$

a non-negative simple function. In words: φ_n rounds f down to the nearest multiple of 2^{-n}, on the region where $f < n$, and sets $\varphi_n = n$ where $f \geq n$.

Monotonicity. Passing from level n to level $n+1$ refines each interval $[k/2^n, (k+1)/2^n)$ into

$$\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) = \left[\frac{2k}{2^{n+1}}, \frac{2k+1}{2^{n+1}}\right) \cup \left[\frac{2k+1}{2^{n+1}}, \frac{2k+2}{2^{n+1}}\right). \tag{2.27}$$

So $A_{k,n} = A_{2k,\,n+1} \cup A_{2k+1,\,n+1}$; on either piece, the assigned value at level $n+1$ is at least the value $k/2^n$ assigned at level n. Likewise, the top set $A_{n2^n,n} = \{f \geq n\}$ splits, at level $n+1$, into the part where $n \leq f < n+1$ (now assigned a value $\geq n$, refined further) and the part where $f \geq n+1$ (assigned the value $n+1 > n$). Hence $\varphi_n \leq \varphi_{n+1}$ pointwise.

Pointwise convergence. Fix $x \in \Omega$. If $f(x) = \infty$, then $\varphi_n(x) = n \to \infty = f(x)$. If $f(x) < \infty$, then once $n > f(x)$, we have $x \in A_{k,n}$ for the unique k with

$$\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}, \tag{2.28}$$

so

$$0 \leq f(x) - \varphi_n(x) < \frac{1}{2^n} \longrightarrow 0. \tag{2.29}$$

Hence $\varphi_n(x) \to f(x)$ for every $x \in \Omega$. ∎

Remark 2.5. The convergence in this proposition holds **everywhere**, not merely almost everywhere, and the approximating sequence is **monotonically increasing**. Both features will be essential in the proof of the Monotone Convergence Theorem below.

§2.iv. The Integral of a Non-negative Measurable Function

Definition 2.4 (Integral of a non-negative measurable function)**.** Let $f : \Omega \to [0, \infty]$ be measurable. We define

$$\int_\Omega f \, \mathrm{d}\mu := \sup \left\{ \int_\Omega \varphi \, \mathrm{d}\mu \ : \ \varphi \text{ simple}, \ 0 \leq \varphi \leq f \right\} \in [0, \infty]. \tag{2.30}$$

Remark 2.6. This definition could have been adopted directly, without first proving the approximation proposition above; the proposition simply confirms that the class of simple functions lying below f is rich enough to be useful (e.g. in proving the Monotone Convergence Theorem).

Notation for the integral varies across the literature and is otherwise a matter of convention: one may write $\int_\Omega f \, \mathrm{d}\mu$, $\int_\Omega f(x) \, \mathrm{d}\mu(x)$, $\int f \, \mathrm{d}\mu$ (when Ω is clear), or $\int f$ (when μ is also clear); some authors (especially in physics) write the measure before the function. Any of these is acceptable as long as the underlying set, σ-algebra, and measure are unambiguous.

§2.v. Basic Properties of the Integral

The following two properties follow immediately from the definition as a supremum.

Proposition 2.4 (Monotonicity)

If $f, g : \Omega \to [0, \infty]$ are measurable and $f \leq g$, then

$$\int_\Omega f \, \mathrm{d}\mu \leq \int_\Omega g \, \mathrm{d}\mu. \tag{2.31}$$

Proof. Every simple function φ with $0 \leq \varphi \leq f$ also satisfies $0 \leq \varphi \leq g$, so the set over which the supremum defining $\int_\Omega f \, \mathrm{d}\mu$ is taken is contained in the set over which the supremum defining $\int_\Omega g \, \mathrm{d}\mu$ is taken. A supremum over a smaller set cannot exceed a supremum over a larger set. ∎

Proposition 2.5 (Positive homogeneity)

If $f : \Omega \to [0, \infty]$ is measurable and $c \geq 0$, then

$$\int_\Omega cf \, \mathrm{d}\mu = c \int_\Omega f \, \mathrm{d}\mu. \tag{2.32}$$

Proof. This conclusion clearly holds if $c = 0$. For $c > 0$, note that $\varphi \leq cf \iff \frac{\varphi}{c} \leq f$. ∎

Remark 2.7. Additivity of the integral, $\int_\Omega (f + g) \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu + \int_\Omega g \, \mathrm{d}\mu$ for non-negative measurable f, g, is *not* immediate from the definition and requires a genuine proof; this will follow once the Monotone Convergence Theorem is available (approximate f and g by increasing sequences of simple functions, for which additivity is straightforward, and pass to the limit).

Definition 2.5 (Integral over a subset)**.** Let $A \in \Sigma$ and let $g : \Omega \to [0, \infty]$ be measurable. The integral of g over A is defined by

$$\int_A g \, \mathrm{d}\mu := \int_\Omega g \, \chi_A \, \mathrm{d}\mu. \tag{2.33}$$

§2.vi. The Monotone Convergence Theorem

We now come to the first of the two fundamental convergence theorems of integration theory (the second being the Dominated Convergence Theorem, to be treated in a later lecture). Both address the same basic question: when can a limit be passed inside an integral?

Theorem 2.6 (Monotone Convergence Theorem)

Let $(f_n)_{n \geq 1}$ be a sequence of measurable functions $\Omega \to [0, \infty]$ such that

$$0 \leq f_n \leq f_{n+1} \qquad \text{for every } n, \tag{2.34}$$

and let $f(x) := \lim_{n \to \infty} f_n(x)$ (which exists in $[0, \infty]$ for every x, by monotonicity). Then f is measurable, and

$$\int_\Omega f \, \mathrm{d}\mu = \lim_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{2.35}$$

Proof. Since $(f_n(x))_n$ is non-decreasing for each x, its limit equals $\limsup_n f_n(x) = \sup_n f_n(x)$, which is measurable by .

For each n, since $f_n \leq f$, monotonicity of the integral gives

$$\int_\Omega f_n \, \mathrm{d}\mu \leq \int_\Omega f \, \mathrm{d}\mu \qquad \text{for all } n. \tag{$\star$}$$

The left-hand side is non-decreasing in n (again by monotonicity, since $f_n \leq f_{n+1}$) and bounded above by the fixed quantity $\int_\Omega f \, \mathrm{d}\mu$, so the limit exists and

$$\lim_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu \;\leq\; \int_\Omega f \, \mathrm{d}\mu. \tag{2.36}$$

Fix a simple function φ with $0 \le \varphi \le f$, and fix a constant $\alpha \in (0,1)$. Define

$$A_n := \{x \in \Omega : f_n(x) \ge \alpha\,\varphi(x)\}. \tag{2.37}$$

Each A_n is measurable (as $f_n - \alpha\varphi$ is measurable).

Claim 2.2: $A_n \subseteq A_{n+1}$ for every n.

Proof. If $f_n(x) \ge \alpha\varphi(x)$, then since $f_{n+1}(x) \ge f_n(x)$, also $f_{n+1}(x) \ge \alpha\varphi(x)$. □

Claim 2.3: $\bigcup_{n=1}^{\infty} A_n = \Omega$.

Proof. Fix $x \in \Omega$. If $\varphi(x) = 0$, then $x \in A_1$ trivially (as $f_1(x) \ge 0 = \alpha\varphi(x)$). If $\varphi(x) > 0$, then

$$\alpha\varphi(x) < \varphi(x) \le f(x). \tag{2.38}$$

Since $f_n(x) \to f(x)$, eventually $f_n(x) > \alpha\varphi(x)$, i.e. $x \in A_n$ for n large. □

By monotonicity and the definition of the integral over a subset,

$$\int_\Omega f_n \,\mathrm{d}\mu \;\ge\; \int_{A_n} f_n \,\mathrm{d}\mu \;\ge\; \int_{A_n} \alpha\varphi \,\mathrm{d}\mu = \alpha \int_{A_n} \varphi \,\mathrm{d}\mu. \tag{2.39}$$

Since φ is simple, the set function $B \mapsto \int_B \varphi \,\mathrm{d}\mu$ is a measure. Since (A_n) is an increasing sequence of measurable sets with $\bigcup_n A_n = \Omega$, by Proposition 1.4,

$$\lim_{n\to\infty} \int_{A_n} \varphi \,\mathrm{d}\mu = \int_\Omega \varphi \,\mathrm{d}\mu. \tag{2.40}$$

Taking $n \to \infty$ in (2.39),

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \;\ge\; \alpha \int_\Omega \varphi \,\mathrm{d}\mu. \tag{2.41}$$

This holds for every $\alpha \in (0,1)$; letting $\alpha \to 1^-$ gives

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \;\ge\; \int_\Omega \varphi \,\mathrm{d}\mu. \tag{2.42}$$

Finally, taking the supremum over all simple φ with $0 \le \varphi \le f$,

$$\lim_{n\to\infty} \int_\Omega f_n \,\mathrm{d}\mu \;\ge\; \int_\Omega f \,\mathrm{d}\mu. \tag{2.43}$$

Combining both inequalities gives the result. ■

Remark 2.8. The restriction $\alpha < 1$ (rather than simply taking $\alpha = 1$, i.e. $A_n = \{f_n \ge \varphi\}$) is essential: without it, claim (b) above can fail. It is possible for a simple function $\varphi \le f$ to touch f exactly at some point x, while the approximating sequence $f_n(x)$ increases to $f(x)$ without ever reaching it, so that $f_n(x) < \varphi(x)$ for every finite n.

Remark 2.9. The Monotone Convergence Theorem is, in essence, already implicit in the construction of $\int_\Omega f \, \mathrm{d}\mu$. The integral of f was defined as a supremum of integrals of simple functions below f, and we exhibited an increasing sequence of simple functions converging to f everywhere. The Monotone Convergence Theorem shows that this phenomenon persists for *any* increasing sequence of non-negative measurable functions converging to f, not only sequences of simple functions.

Lemma 2.7

If $f_1, f_2 \in \mathcal{M}^+$, then $\int_\Omega (f_1 + f_2) \, \mathrm{d}\mu = \int_\Omega f_1 \, \mathrm{d}\mu + \int_\Omega f_2 \, \mathrm{d}\mu$. By induction,

$$\int_\Omega \sum_{i=1}^{n} f_i \, \mathrm{d}\mu = \sum_{i=1}^{n} \int_\Omega f_i \, \mathrm{d}\mu. \tag{2.44}$$

Proof. By Theorem 2.3, we can find increasing sequence of simple functions $\varphi_n \to f_1, \psi_n \to f_2$. Then $\varphi_n + \psi_n$ is an increasing sequence of simple function that monotonically converges to $f_1 + f_2$. By Monotone Convergence Theorem,

$$\begin{aligned} \int_\Omega (f_1 + f_2) \, \mathrm{d}\mu &= \lim_{n\to\infty} \int_\Omega (\varphi_n + \psi_n) \, \mathrm{d}\mu && (2.45)\\ &= \lim_{n\to\infty} \int_\Omega \varphi_n \, \mathrm{d}\mu + \lim_{n\to\infty} \int_\Omega \psi_n \, \mathrm{d}\mu && (2.46)\\ &= \int_\Omega f_1 \, \mathrm{d}\mu + \int_\Omega f_2 \, \mathrm{d}\mu. && (2.47) \end{aligned}$$

■

Proposition 2.8

If $\{f_n\}_n$ is a sequence in $\mathcal{M}^+$, then

$$\int_\Omega \sum_{i=1}^{\infty} f_i \, \mathrm{d}\mu = \sum_{i=1}^{\infty} \int_\Omega f_i \, \mathrm{d}\mu. \tag{2.48}$$

Proof. $g_n = \sum_{i=1}^{n} f_n$ is a monotone increasing sequence converging to $f = \sum_{i=1}^{\infty} f_i$. Therefore, by Monotone Convergence Theorem,

$$\int_\Omega \sum_{i=1}^{\infty} f_i \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega g_n \, \mathrm{d}\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int_\Omega f_i \, \mathrm{d}\mu = \sum_{i=1}^{\infty} \int_\Omega f_i \, \mathrm{d}\mu. \tag{2.49}$$

■

Proposition 2.9

If $f \in \mathcal{M}^+$ and $\int_\Omega f \, \mathrm{d}\mu = 0$, then $f = 0$ μ-a.e.

Proof. Suppose $f = 0$ a.e. Let E be the set where f is nonzero, then $\mu(E) = 0$ by hypothesis. Then consider a simple function in its standard form $\varphi = \sum_{i=1}^{n} a_i \chi_{A_i} \leq f$, with $a_i > 0$ and A_i's pairwise disjoint. Since $\varphi \leq f$, hence each $A_i \subseteq E$, and consequently, $\mu(A_i) = 0$. Then

$$\int_\Omega \varphi \, \mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(A_i) = 0 \implies \int_\Omega f \, \mathrm{d}\mu = \sup_{\substack{0 \leq \varphi \leq f \\ \varphi \text{ simple}}} \int_\Omega \varphi \, \mathrm{d}\mu = 0. \tag{2.50}$$

Conversely, suppose $\int_\Omega f \, \mathrm{d}\mu = 0$. Consider the set

$$E_n = \left\{ x \in \Omega : f(x) > \frac{1}{n} \right\}. \tag{2.51}$$

Then we have

$$E := \{ x \in \Omega : f(x) > 0 \} = \bigcup_{n=1}^{\infty} E_n. \tag{2.52}$$

We need to show that $\mu(E) = 0$. Assume for the sake of contradiction that it's strictly greater than 0. By continuity (Proposition 1.4),

$$0 < \mu(E) = \lim_{n \to \infty} \mu(E_n), \tag{2.53}$$

implies that there exists some m for which $\mu(E_m) > 0$. Then $0 \leq \frac{1}{m} \chi_{E_m} \leq f$, but then

$$\int_\Omega f \, \mathrm{d}\mu \geq \int_\Omega \frac{1}{m} \chi_{E_m} \, \mathrm{d}\mu = \frac{1}{m} \mu(E_m) > 0. \tag{2.54}$$

■

Theorem 2.10

If $\{f_n\}_{n \geq 1}$ and f are functions in $\mathcal{M}^+$ such that $\{f_n\}_{n \geq 1}$ converges monotonically increasing to f for μ-a.e. $x \in \Omega$. Then conclusion of Monotone Convergence Theorem still holds, i.e.

$$\int_\Omega f \, \mathrm{d}\mu = \lim_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{2.55}$$

Proof. Suppose $f_n(x) \to f(x)$ monotonically increasingly on $E \subseteq \Omega$, then $\mu(E^c) = 0$. Then $f - f\chi_E = 0$ and $f_n - f_n \chi_E = 0$ a.e. Also, $f_n \chi_E$ converges monotonically increasingly to $f\chi_E$ everywhere. As a result, by Monotone Convergence Theorem,

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f\chi_E \, \mathrm{d}\mu = \lim_{n \to \infty} \int_\Omega f_n \chi_E \, \mathrm{d}\mu = \lim_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{2.56}$$

■

Proposition 2.11

If $f \in \mathcal{M}^+$, then the map

$$A \mapsto \int_A f \, \mathrm{d}\mu = \int_\Omega f \cdot \chi_A \, \mathrm{d}\mu \tag{2.57}$$

is a measure on Σ.

Proof. Suppose $\nu(A) = \int_\Omega f \cdot \chi_A \, \mathrm{d}\mu$. Clearly, $\nu(A) \geq 0$. Also

$$\nu(\varnothing) = \int_\Omega f \cdot \chi_\varnothing \, \mathrm{d}\mu = \int_\Omega 0 \, \mathrm{d}\mu = 0. \tag{2.58}$$

Now, suppose $\{A_n\}_{n\geq 1}$ is a countable collection of pairwise disjoint sets. We consider

$$f_n = \sum_{i=1}^{n} (f\chi_{A_i}) = f\chi_{\bigcup_{i=1}^n A_i}, \tag{2.59}$$

which is a monotonically increasing sequence of measurable functions, that converge pointwise to $f\chi_{\bigcup_{i=1}^\infty A_i}$. Then by Monotone Convergence Theorem,

$$\begin{aligned}
\nu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \int_\Omega f\chi_{\bigcup_{i=1}^\infty A_i} \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu && (2.60)\\
&= \lim_{n\to\infty} \int_\Omega \sum_{i=1}^{n} (f\chi_{A_i}) \, \mathrm{d}\mu = \lim_{n\to\infty} \sum_{i=1}^{n} \int_\Omega (f\chi_{A_i}) \, \mathrm{d}\mu && (2.61)\\
&= \sum_{i=1}^{\infty} \nu(A_i). && (2.62)
\end{aligned}$$

■

Corollary 2.12

If $A_1 \subseteq A_2 \subseteq \cdots$ is an increasing sequence of measurable sets and $f \in \mathcal{M}^+$, then

$$\int_{\bigcup_{i=1}^\infty A_i} f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} f \, \mathrm{d}\mu. \tag{2.63}$$

Proof. $A \mapsto \nu(A) = \int_A f \, \mathrm{d}\mu$ is a measure on Σ, by Proposition 2.11. $\{A_n\}_n$ is an increasing sequence of measurable sets. Therefore, by Proposition 1.4,

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \nu(A_n) \implies \int_{\bigcup_{i=1}^\infty A_i} f \, \mathrm{d}\mu = \lim_{n\to\infty} \int_{A_n} f \, \mathrm{d}\mu. \tag{2.64}$$

■

Proposition 2.13

For the measure $A \mapsto \nu(A) = \int_A f \, d\mu$, if $g \in \mathcal{M}^+$,

$$\int_\Omega g \, d\nu = \int_\Omega fg \, d\mu. \tag{2.65}$$

Proof. If $g = \chi_E$ for a measurable set E, then

$$\int_\Omega g \, d\nu = \int_\Omega \chi_E \, d\nu = \nu(E) = \int_E f \, d\mu = \int_\Omega f\chi_E \, d\mu. \tag{2.66}$$

Therefore, (2.65) holds when g is a non-negative simple function. Now, for arbitrary $g \in \mathcal{M}^+$, we can find an increasing sequence of non-negative simple functions

$$0 \leq g_1 \leq g_2 \leq \cdots \tag{2.67}$$

such that $g_n \to g$ pointwise everywhere. Then by Monotone Convergence Theorem,

$$\int_\Omega g \, d\nu = \lim_{n\to\infty} \int_\Omega g_n \, d\nu = \lim_{n\to\infty} \int_\Omega f g_n \, d\mu. \tag{2.68}$$

Since $f \geq 0$, fg_n is an increasing sequence of non-negative functions such that $fg_n \to fg$ pointwise everywhere. Therefore, by Monotone Convergence Theorem,

$$\lim_{n\to\infty} \int_\Omega f g_n \, d\mu = \int_\Omega fg \, d\mu. \tag{2.69}$$

■

Proposition 2.14

If μ and ν are two measures on a measurable space (Ω, Σ), then $\eta = \mu + \nu$, is also a measure on (Ω, Σ). Furthermore, for $f \in \mathcal{M}^+$,

$$\int_\Omega f \, d\eta = \int_\Omega f \, d\mu + \int_\Omega f \, d\nu. \tag{2.70}$$

Proof. Clearly,

$$\eta(\varnothing) = \mu(\varnothing) + \nu(\varnothing) = 0 + 0 = 0. \tag{2.71}$$

If $\{E_n\}_{n=1}^\infty$ is a collection of pairwise disjoint sets

$$\eta\left(\bigcup_{n=1}^\infty E_n\right) = \mu\left(\bigcup_{n=1}^\infty E_n\right) + \nu\left(\bigcup_{n=1}^\infty E_n\right) \tag{2.72}$$

$$= \sum_{n=1}^\infty \mu(E_n) + \sum_{n=1}^\infty \nu(E_n) \tag{2.73}$$

$$= \sum_{n=1}^\infty (\mu+\nu)(E_n) = \sum_{n=1}^\infty \eta(E_n). \tag{2.74}$$

So ν is a measure. Now, if E is a measurable set,

$$\int_\Omega \chi_E \, \mathrm{d}\eta = \eta(E) = \mu(E) + \nu(E) = \int_\Omega \chi_E \, \mathrm{d}\mu + \int_\Omega \chi_E \, \mathrm{d}\nu. \tag{2.75}$$

Hence, (2.70) holds holds when f is a non-negative simple function. Now, for arbitrary $f \in \mathcal{M}^+$, we can find an increasing sequence of non-negative simple functions

$$0 \leq f_1 \leq f_2 \leq \cdots \tag{2.76}$$

such that $f_n \to f$ pointwise everywhere. Then by Monotone Convergence Theorem,

$$\int_\Omega f \, \mathrm{d}\eta = \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\eta = \lim_{n\to\infty} \left(\int_\Omega f_n \, \mathrm{d}\mu + \int_\Omega f_n \, \mathrm{d}\nu \right) \tag{2.77}$$

$$= \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu + \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\nu. \tag{2.78}$$

Now, since f_n is an increasing sequence of non-negative functions such that $f_n \to f$ pointwise everywhere, by Monotone Convergence Theorem,

$$\lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\mu = \int_\Omega f \, \mathrm{d}\mu \text{ and } \lim_{n\to\infty} \int_\Omega f_n \, \mathrm{d}\nu = \int_\Omega f \, \mathrm{d}\nu. \tag{2.79}$$

Plugging these into (2.78), we are done! ■

§3. 14 September 2026

In our last lecture, we proved that if f_n is a monotonically increasing sequence of measurable functions, then we can essentially swap limit and integral:

$$\int_\Omega \lim_{n\to\infty} f_n \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu. \tag{3.1}$$

The hypothesis that (f_n) is an *increasing* sequence is essential, as the following examples show.

Example 3.1

Consider $f_n = \chi_{(n,n+1)}$. Then Then $f_n(x) \to 0$ pointwise for every $x \in \mathbb{R}$, yet

$$\int_{\mathbb{R}} f_n \, d\mu = 1 \quad \text{for every } n, \qquad \text{while} \qquad \int_{\mathbb{R}} \left(\lim_{n\to\infty} f_n\right) d\mu = 0. \tag{3.2}$$

So the conclusion of Monotone Convergence Theorem fails without monotonicity.

Example 3.2

Define, for $n \geq 1$,

$$f_n = n\,\chi_{\left(0,\frac{1}{n}\right)}, \tag{3.3}$$

or, to make it continuous, a triangular spike on $[0, 1/n]$ of height $2n$ so that its graph is a triangle of area 1. Then

$$\int_0^1 f_n(x)\, dx = 1 \quad \text{for every } n, \tag{3.4}$$

but for every fixed $x > 0$ there is N such that $1/n < x$ for all $n \geq N$, so $f_n(x) \to 0$ pointwise (and $f_n(0) = 0$ for all n, by construction). Hence $f_n \to 0$ pointwise while $\int_0^1 f_n \, d\mu = 1$, again showing that monotonicity cannot be dropped.

§3.i. Fatou's lemma

This type of example is the classical illustration that pointwise convergence alone does not let one pass the limit inside the integral; one really needs either monotonicity or a dominating integrable function (Dominated Convergence Theorem, see below).

Theorem 3.1 (Fatou's Lemma)

Let (f_n) be a sequence of nonnegative measurable functions on Ω. Then

$$\int_\Omega \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_\Omega f_n \, d\mu. \tag{3.5}$$

Proof. For each m, define

$$g_m(x) = \inf_{n\geq m} f_n(x). \tag{3.6}$$

For every $n \geq m$ we have $g_m \leq f_n$, and since the integral is monotone,

$$\int g_m \, \mathrm{d}\mu \leq \int f_n \, \mathrm{d}\mu \qquad \text{for all } n \geq m. \tag{3.7}$$

Taking the infimum over $n \geq m$ on the right-hand side gives

$$\int g_m \, \mathrm{d}\mu \leq \inf_{n \geq m} \int f_n \, \mathrm{d}\mu. \tag{3.8}$$

The sequence (g_m) is increasing in m (an infimum over a smaller index set is larger), and

$$\lim_{m \to \infty} g_m(x) = \sup_m \inf_{n \geq m} f_n(x) = \liminf_{n \to \infty} f_n(x). \tag{3.9}$$

By the Monotone Convergence Theorem applied to (g_m),

$$\int \left(\liminf_{n \to \infty} f_n \right) \mathrm{d}\mu = \lim_{m \to \infty} \int g_m \, \mathrm{d}\mu \;\leq\; \lim_{m \to \infty} \inf_{n \geq m} \int f_n \, \mathrm{d}\mu = \liminf_{n \to \infty} \int f_n \, \mathrm{d}\mu. \tag{3.10}$$

■

Proposition 3.2

If $\{f_n\}_{n \geq 1}$ and f are functions in $\mathcal{M}^+$ such that $f_n(x)$ converges to $f(x)$ for μ-a.e. $x \in \Omega$, then the conclusion of Fatou's Lemma still holds, i.e.

$$\int_\Omega f \, \mathrm{d}\mu \leq \liminf_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.11}$$

Proof. Suppose $f_n(x) \to f(x)$ on $E \subseteq \Omega$, then $\mu(E^c) = 0$. Then $f - f\chi_E = 0$ and $f_n - f_n\chi_E = 0$ a.e. Also, $f_n\chi_E$ converges to $f\chi_E$ everywhere. As a result,

$$\int_\Omega f \, \mathrm{d}\mu = \int_\Omega f\chi_E \, \mathrm{d}\mu \leq \liminf_{n \to \infty} \int_\Omega f_n\chi_E \, \mathrm{d}\mu = \liminf_{n \to \infty} \int_\Omega f_n \, \mathrm{d}\mu. \tag{3.12}$$

■

Proposition 3.3

If $f \in \mathcal{M}^+$ and $\int_\Omega f \, \mathrm{d}\mu < \infty$, then $\{x \in \Omega : f(x) = \infty\}$ is a null set and $\{x \in \Omega : f(x) > 0\}$ is σ-finite.

Proof. Suppose $E := \{x \in \Omega : f(x) = \infty\}$. Assume for the sake of contradiction that $\mu(E) > 0$. Then $f \geq f \cdot \chi_E$. So

$$\int_\Omega f \, \mathrm{d}\mu \geq \int_\Omega f \cdot \chi_E \, \mathrm{d}\mu = \int_\Omega \infty \cdot \chi_E \, \mathrm{d}\mu = \infty \cdot \mu(E) = \infty, \tag{3.13}$$

which is a contradiction! Now, suppose $F = \{x \in \Omega : f(x) > 0\}$. Observe that

$$F = \bigcup_{n=1}^{\infty} \left\{ x \in \Omega : f(x) \geq \frac{1}{n} \right\} =: \bigcup_{n=1}^{\infty} F_n. \tag{3.14}$$

So it suffices to prove that each F_n has finite measure. Note that $f(x) \geq \frac{1}{n}$ on F_n, so that $n \cdot f(x) \geq 1$. In other words, $\chi_{F_n} \leq n \cdot f$. By monotonicity of integrals,

$$\mu(F_n) = \int_\Omega \chi_{F_n} \, \mathrm{d}\mu \leq \int_\Omega n \cdot f \, \mathrm{d}\mu = n \int_\Omega f \, \mathrm{d}\mu < \infty. \tag{3.15}$$

■

§3.ii. Integrable functions

Definition 3.1 (Positive and negative part)**.** For a measurable function $f : \Omega \to \mathbb{R}$ define

$$f^+(x) = \max\{f(x), 0\}, \qquad f^-(x) = \max\{-f(x), 0\}, \tag{3.16}$$

to be the **positive part** and **negative part** of f, respectively.

Remark 3.1. It's clear that $f^+, f^- \geq 0$ are measurable, and

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-. \tag{3.17}$$

This decomposition is not unique as a way of writing f as a difference of two nonnegative functions (one may add the same nonnegative function to both f^+ and f^-), but it is the canonical choice, and it is the one used to define the integral of a general function.

Definition 3.2 (The space L^1)**.** Let (Ω, μ) be a measure space. Define

$$L^1(\Omega, \mu) = \left\{ f : \Omega \to \mathbb{R} \text{ measurable } : \int_\Omega |f| \, \mathrm{d}\mu < \infty \right\}. \tag{3.18}$$

For $f \in L^1(\Omega, \mu)$, both $\int f^+ \, \mathrm{d}\mu$ and $\int f^- \, \mathrm{d}\mu$ are finite (since their sum is $\int |f| \, \mathrm{d}\mu < \infty$), and we define

$$\int_\Omega f \, \mathrm{d}\mu := \int_\Omega f^+ \, \mathrm{d}\mu - \int_\Omega f^- \, \mathrm{d}\mu. \tag{3.19}$$

Remark 3.2. The integral is first defined for simple functions, then extended to nonnegative measurable functions via the supremum definition, and finally extended to L^1 via this difference, precisely so as to avoid ever having to make sense of $\infty - \infty$.

Proposition 3.4

If $f \in L^1$, then $\left| \int_\Omega f \, \mathrm{d}\mu \right| \leq \int_\Omega |f| \, \mathrm{d}\mu$.

Proof.

$$\left|\int_\Omega f\,\mathrm{d}\mu\right| = \left|\int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu\right| \leq \int_\Omega f^+\,\mathrm{d}\mu + \int_\Omega f^-\,\mathrm{d}\mu = \int_\Omega |f|\,\mathrm{d}\mu. \tag{3.20}$$

■

Proposition 3.5

If $f, g \in L^1(\Omega, \mu)$ and $c \in \mathbb{R}$, then $f + cg \in L^1(\Omega, \mu)$ and

$$\int_\Omega (f + cg)\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + c\int_\Omega g\,\mathrm{d}\mu. \tag{3.21}$$

Proof. First notice that

$$|f + cg| \leq |f| + |c|\,|g|\,, \tag{3.22}$$

which implies using monotonicity of integral of non-negative functions that

$$\int_\Omega |f + cg|\,\mathrm{d}\mu \leq \int_\Omega |f| + |c|\,|g|\,\mathrm{d}\mu = \int_\Omega |f|\,\mathrm{d}\mu + |c|\int_\Omega |g|\,\mathrm{d}\mu < \infty. \tag{3.23}$$

We'll now show that $\int_\Omega cf\,\mathrm{d}\mu = c\int_\Omega f\,\mathrm{d}\mu$. For $c \geq 0$,

$$(cf)^+ = cf^+, (cf)^- = cf^-. \tag{3.24}$$

Then

$$\int_\Omega cf\,\mathrm{d}\mu = \int_\Omega (cf)^+\,\mathrm{d}\mu - \int_\Omega (cf)^-\,\mathrm{d}\mu = \int_\Omega cf^+\,\mathrm{d}\mu - \int_\Omega cf^-\,\mathrm{d}\mu = c\int_\Omega f\,\mathrm{d}\mu. \tag{3.25}$$

For $c < 0$,

$$(cf)^+ = -cf^-, (cf)^- = -cf^+. \tag{3.26}$$

Then

$$\int_\Omega cf\,\mathrm{d}\mu = \int_\Omega (cf)^+\,\mathrm{d}\mu - \int_\Omega (cf)^-\,\mathrm{d}\mu = \int_\Omega -cf^-\,\mathrm{d}\mu - \int_\Omega -cf^+\,\mathrm{d}\mu = c\int_\Omega f\,\mathrm{d}\mu. \tag{3.27}$$

Now let's show that $\int_\Omega f + g\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu$. Suppose $h = f + g$. Then

$$h^+ - h^- = f^+ - f^- + g^+ - g^- \implies h^+ + f^- + g^- = h^- + f^+ + g^+. \tag{3.28}$$

Using the additivity of integral of non-negative functions, we get

$$\int_\Omega h^+\,\mathrm{d}\mu + \int_\Omega f^-\,\mathrm{d}\mu + \int_\Omega g^-\,\mathrm{d}\mu = \int_\Omega h^-\,\mathrm{d}\mu + \int_\Omega f^+\,\mathrm{d}\mu + \int_\Omega g^+\,\mathrm{d}\mu. \tag{3.29}$$

Regrouping gives us the desired result:

$$\int_\Omega h^+\,\mathrm{d}\mu - \int_\Omega h^-\,\mathrm{d}\mu = +\int_\Omega f^+\,\mathrm{d}\mu - \int_\Omega f^-\,\mathrm{d}\mu + \int_\Omega g^+\,\mathrm{d}\mu - \int_\Omega g^-\,\mathrm{d}\mu, \tag{3.30}$$

i.e. $\int_\Omega h\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu + \int_\Omega g\,\mathrm{d}\mu$. ■

Lemma 3.6

If $f \le g$ a.e. then $\int_\Omega f\,\mathrm{d}\mu \le \int_\Omega g\,\mathrm{d}\mu$.

Proof. Suppose $f \le g$ on E with $\mu(E^c) = 0$. Then $(g-f)\chi_E \ge 0$. The integral of a non-negative function is non-negative, so

$$\int_\Omega g\,\mathrm{d}\mu - \int_\Omega f\,\mathrm{d}\mu = \int_\Omega (g-f)\,\mathrm{d}\mu = \int_\Omega (g-f)\chi_E\,\mathrm{d}\mu \ge 0. \tag{3.31}$$

■

Proposition 3.7

If $f \in M$ and $g \in L^1(\Omega,\mu)$ with $|f| \le g$ a.e., then $f \in L^1(\Omega,\mu)$. (A function dominated by an integrable function is itself integrable.)

Proof. By monotonicity of integrals,

$$\int_\Omega |f|\,\mathrm{d}\mu \le \int_\Omega g\,\mathrm{d}\mu < \infty. \tag{3.32}$$

■

Proposition 3.8

If $f, g \in L^1$, then the following are equivalent:

(a) $\int_A f\,\mathrm{d}\mu = \int_A g\,\mathrm{d}\mu$ for every $A \in \Sigma$;

(b) $\int_\Omega |f-g|\,\mathrm{d}\mu = 0$;

(c) $f = g$ a.e.

Proof. The equivalence of (b) and (c) follows from Proposition 2.9:

$$\int_\Omega |f-g|\,\mathrm{d}\mu = 0 \iff |f-g| = 0 \text{ a.e.} \iff f = g \text{ a.e.} \tag{3.33}$$

(b)⇒(a) Take any $A \in \Sigma$. Then

$$\left|\int_A f\,\mathrm{d}\mu - \int_A g\,\mathrm{d}\mu\right| = \left|\int_\Omega (f-g)\chi_A\,\mathrm{d}\mu\right| \le \int_\Omega |f-g|\,\chi_A\,\mathrm{d}\mu \le \int_\Omega |f-g|\,\mathrm{d}\mu = 0. \tag{3.34}$$

(a)⇒(c) Suppose $A = \{x \in \Omega : f(x) > g(x)\}$, which is a measurable set. On A, $f - g > 0$. If $\mu(A) > 0$, then $\int_A (f-g) > 0$ by the following claim, which contradicts with the hypothesis.

Claim 3.1: If $\mu(E) > 0$ and $h > 0$ on E, then $\int_E h \, d\mu > 0$.

Proof. Let $E_n = \left\{x \in E : h(x) \geq \frac{1}{n}\right\}$. Then $\{E_n\}_n$ is an increasing sequence with

$$E = \bigcup_{i=1}^{\infty} E_n. \tag{3.35}$$

By continuity (Proposition 1.4),

$$0 < \mu(E) = \lim_{n\to\infty} \mu(E_n), \tag{3.36}$$

implies that there exists some m for which $\mu(E_m) > 0$. Then $0 \leq \frac{1}{m}\chi_{E_m} \leq h$, but then

$$\int_E h \, d\mu \geq \int_\Omega \frac{1}{m}\chi_{E_m} \, d\mu = \frac{1}{m}\mu(E_m) > 0. \tag{3.37}$$

□

Therefore, $\mu(A) = 0$. Similarly, if $B = \{x \in \Omega : f(x) < g(x)\}$, then $\mu(B) = 0$. Now,

$$\{x \in \Omega : f(x) \neq g(x)\} = A \cup B, \tag{3.38}$$

is also a null set, since it's the union of two null sets. ■

Remark 3.3. Note that $\int_\Omega f \, d\mu = \int_\Omega g \, d\mu$ does not necessarily imply that $f = g$ a.e. For example, consider $\Omega = [0, 2\pi]$ and $f(x) = \sin x, g(x) = \cos x$. Then

$$\int_0^{2\pi} \sin x \, dx = \int_0^{2\pi} \cos x \, dx = 0, \tag{3.39}$$

but it's **NOT** true that $\sin x = \cos x$ a.e.

§3.iii. Dominated convergence theorem

Theorem 3.9 (Dominated Convergence Theorem)

Let (f_n) be a sequence of measurable functions on Ω such that $f_n \to f$ pointwise a.e., and suppose there is a function $g \in L^1(\Omega, \mu)$ such that

$$|f_n(x)| \leq g(x) \text{ a.e.} \qquad \text{for all } n. \tag{3.40}$$

Then $f \in L^1(\Omega, \mu)$ and

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega f_n \, d\mu. \tag{3.41}$$

Proof. Since $|f_n| \leq g$ for every n and $f_n \to f$ pointwise, also $|f| \leq g$, so $f \in L^1(\Omega, \mu)$ Proposition 3.7.

Since $g + f_n \geq 0$ (because $|f_n| \leq g$) and $g + f_n \to g + f$ pointwise a.e., Fatou's Lemma gives

$$\int_\Omega (g+f)\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega (g+f_n)\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu + \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu. \tag{3.42}$$

Since $\int_\Omega g\,\mathrm{d}\mu$ is finite, it may be cancelled from both sides:

$$\int_\Omega f\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu. \tag{3.43}$$

Likewise $g - f_n \geq 0$ and $g - f_n \to g - f$ pointwise a.e., so Fatou's Lemma gives

$$\int_\Omega (g-f)\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega (g-f_n)\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu - \limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu, \tag{3.44}$$

using $\liminf(-a_n) = -\limsup(a_n)$. Cancelling $\int_\Omega g\,\mathrm{d}\mu$ (finite) from both sides:

$$\limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu \leq \int_\Omega f\,\mathrm{d}\mu. \tag{3.45}$$

Combining (3.43) and (3.45),

$$\limsup_{n\to\infty} \int f_n\,\mathrm{d}\mu \leq \int f\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n\,\mathrm{d}\mu, \tag{3.46}$$

and since always $\liminf \leq \limsup$, all three quantities coincide. Hence $\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu$ exists and equals $\int_\Omega f\,\mathrm{d}\mu$. ∎

Remark 3.4. The Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem are the three fundamental convergence theorems needed to justify moving a limit inside an integral: either the sequence is monotone (Monotone Convergence), or it is dominated by a single integrable function (Dominated Convergence); Fatou's Lemma always applies to nonnegative sequences and is the tool used to prove the other two.

Theorem 3.10 (Generalized Dominated Convergence Theorem)

If $f_n, g_n, g \in L^1$, $f_n \to f$ and $g_n \to g$ a.e.,

$$|f_n| \leq g_n, \quad \text{and} \quad \lim_{n\to\infty} \int_\Omega g_n\,\mathrm{d}\mu = \int_\Omega g\,\mathrm{d}\mu, \tag{3.47}$$

then $f \in L^1$, and

$$\lim_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu = \int_\Omega f\,\mathrm{d}\mu. \tag{3.48}$$

Proof. Since $|f_n| \leq g_n$ for every n and $f_n \to f$ pointwise a.e. and $g_n \to g$ a.e., we have that $|f| \leq g$ a.e. So $f \in L^1(\Omega, \mu)$ Proposition 3.7.

Since $g_n + f_n \geq 0$ (because $|f_n| \leq g_n$) and $g_n + f_n \to g + f$ pointwise a.e., Fatou's Lemma gives

$$\begin{aligned}\int_\Omega g\,\mathrm{d}\mu + \int_\Omega f\,\mathrm{d}\mu &= \int_\Omega (g+f)\,\mathrm{d}\mu \\ &\leq \liminf_{n\to\infty} \int_\Omega (g_n + f_n)\,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n\,\mathrm{d}\mu + \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu \\ &= \int_\Omega g\,\mathrm{d}\mu + \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu. \end{aligned} \tag{3.49}$$

Since $\int_\Omega g\,\mathrm{d}\mu$ is finite, it may be cancelled from both sides:

$$\int_\Omega f\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu. \tag{3.50}$$

Likewise $g_n - f_n \geq 0$ and $g_n - f_n \to g - f$ pointwise a.e., so Fatou's Lemma gives

$$\begin{aligned}\int_\Omega g\,\mathrm{d}\mu - \int_\Omega f\,\mathrm{d}\mu &= \int_\Omega (g-f)\,\mathrm{d}\mu \\ &\leq \liminf_{n\to\infty} \int_\Omega (g_n - f_n)\,\mathrm{d}\mu \\ &= \lim_{n\to\infty} \int_\Omega g_n\,\mathrm{d}\mu - \limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu \\ &= \int_\Omega g\,\mathrm{d}\mu - \limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu. \end{aligned} \tag{3.51}$$

Cancelling $\int_\Omega g\,\mathrm{d}\mu$ (finite) from both sides:

$$\limsup_{n\to\infty} \int_\Omega f_n\,\mathrm{d}\mu \leq \int_\Omega f\,\mathrm{d}\mu. \tag{3.52}$$

Combining (3.50) and (3.52),

$$\limsup_{n\to\infty} \int f_n\,\mathrm{d}\mu \leq \int f\,\mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n\,\mathrm{d}\mu, \tag{3.53}$$

and since always $\liminf \leq \limsup$, all three quantities coincide. Hence $\lim_{n\to\infty} \int f_n\,\mathrm{d}\mu$ exists and equals $\int_\Omega f\,\mathrm{d}\mu$. ∎

Remark 3.5. In the proof of Generalized Dominated Convergence Theorem, we used the fact that if a_n, b_n are two sequences such that $a_n \to a$, then

$$\liminf_{n\to\infty} (a_n + b_n) = a + \liminf_{n\to\infty} b_n, \tag{3.54}$$

$$\liminf_{n\to\infty} (a_n - b_n) = a - \limsup_{n\to\infty} b_n. \tag{3.55}$$

To prove the first one, notice that

$$\inf_{m\geq n} (a_m + b_m) \geq \inf_{m\geq n} a_m + \inf_{m\geq n} b_m. \tag{3.56}$$

Taking $n \to \infty$, we get

$$\liminf_{n\to\infty} (a_n + b_n) \geq \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \tag{3.57}$$

$$= a + \liminf_{n\to\infty} b_n. \tag{3.58}$$

On the other hand, writing $b_n = (a_n + b_n) + (-a_n)$ and applying (3.57), we get

$$\liminf_{n\to\infty} b_n = \liminf_{n\to\infty} [(a_n + b_n) + (-a_n)] \tag{3.59}$$

$$\geq \liminf_{n\to\infty} (a_n + b_n) + \liminf_{n\to\infty} (-a_n) \tag{3.60}$$

$$= \liminf_{n\to\infty} (a_n + b_n) - a. \tag{3.61}$$

Therefore, combining (3.58) and (3.61), we have established (3.54). To see (3.55),

$$\liminf_{n\to\infty} (a_n - b_n) = a + \liminf_{n\to\infty} (-b_n) = a - \limsup_{n\to\infty} b_n.$$

Example 3.3

Let

$$f_n(x) = n\,\chi_{(0,1/n)}(x), \qquad x \in [0,1]. \tag{3.62}$$

Then $f_n(x) \to 0$ for every $x \in [0,1]$ (for $x = 0$ every $f_n(0) = 0$; for $x > 0$, eventually $\frac{1}{n} < x$), so $f_n \to 0$ pointwise. However,

$$\int_0^1 f_n \,\mathrm{d}x = 1 \quad \text{for every } n, \qquad \text{while} \int_0^1 \Big(\lim_{n\to\infty} f_n\Big)\,\mathrm{d}x = 0. \tag{3.63}$$

The conclusion of the Dominated Convergence Theorem fails here because no single integrable function g dominates all the f_n. The pointwise supremum of the family is essentially $g(x) = \frac{1}{x}$, and

$$\int_0^1 \frac{1}{x}\,\mathrm{d}x = \infty, \tag{3.64}$$

so $g \notin L^1([0,1])$. This also shows the Monotone Convergence Theorem does not apply, since the sequence (f_n) is not pointwise increasing.

§3.iv. A non-measurable set

We now show that one cannot hope to define a measure on *all* subsets of $\mathbb{R}$ that

(i) is translation invariant,

(ii) is monotone (if $A \subseteq B$ then $\mu(A) \leq \mu(B)$), and

(iii) and that extends the usual length of intervals.

Consequently the σ-algebra of measurable sets must be a strict subset of the power set $\mathcal{P}(\mathbb{R})$. For this, we illustrate the following example:

On $[0,1]$, declare $x \sim y$ if $x - y \in \mathbb{Q}$. This is an equivalence relation, partitioning $[0,1]$ into equivalence classes. By the Axiom of Choice, choose exactly one representative x_I from each equivalence class I, and let

$$A = \{x_I : I \in [0,1]/\!\sim\} \subseteq [0,1], \tag{3.65}$$

where I indexes the set of equivalence classes. The set A is called a *Vitali set*.

Theorem 3.11

There is no measure μ on $\mathcal{P}(\mathbb{R})$ that is translation invariant, monotone, and assigns to every interval its usual length.

Proof. Suppose such a μ exists. For $r \in \mathbb{Q} \cap [-1,1]$ set

$$Y_r = A + r = \{a + r : a \in A\}. \tag{3.66}$$

If $Y_r \cap Y_s \neq \varnothing$ for $r \neq s \in \mathbb{Q}$, then $x_i + r = x_j + s$ for some representatives, so

$$x_i - x_j = s - r \in \mathbb{Q}, \tag{3.67}$$

forcing $x_i \sim x_j$; since A contains only one representative per class, $x_i = x_j$, hence $r = s$, a contradiction. So the sets $\{Y_r\}_{r \in \mathbb{Q} \cap [-1,1]}$ are pairwise disjoint.

Since $A \subseteq [0,1]$ and $r \in [-1,1]$,

$$Y_r \subseteq [-1,2] \qquad \text{for every } r \in \mathbb{Q} \cap [-1,1]. \tag{3.68}$$

For any $x \in [0,1]$ there is a representative $x_i \sim x$ with $x - x_i \in \mathbb{Q} \cap [-1,1]$, so $x \in Y_{x - x_i}$. Hence

$$[0,1] \;\subseteq\; \bigcup_{r \in \mathbb{Q} \cap [-1,1]} Y_r \;\subseteq\; [-1,2]. \tag{3.69}$$

By translation invariance, $\mu(Y_r) = \mu(A)$ for every r. By monotonicity and countable additivity on the disjoint countable union,

$$\mu([0,1]) \;\leq\; \mu\left(\bigcup_{r \in \mathbb{Q} \cap [-1,1]} Y_r \right) = \sum_{r \in \mathbb{Q} \cap [-1,1]} \mu(Y_r) = \sum_{r \in \mathbb{Q} \cap [-1,1]} \mu(A) \;\leq\; \mu([-1,2]). \tag{3.70}$$

Now $\mu([0,1]) = 1$ and $\mu([-1,2]) = 3$ (extending the usual length), so

$$1 \;\leq\; \sum_{r \in \mathbb{Q} \cap [-1,1]} \mu(A) \;\leq\; 3. \tag{3.71}$$

But $\mathbb{Q} \cap [-1,1]$ is countably infinite, so the middle sum is a countable sum of a single fixed value $\mu(A) \in [0, \infty]$:

- if $\mu(A) = 0$, the sum is 0, contradicting $1 \leq 0$;

- if $\mu(A) > 0$, the sum is $+\infty$, contradicting $\infty \leq 3$.

Either way we reach a contradiction. Hence no such measure μ can exist on all of $\mathcal{P}(\mathbb{R})$. ■

Remark 3.6. This is the price one pays for wanting a translation-invariant, monotone measure extending length: it cannot be defined on *every* subset of $\mathbb{R}$. This motivates restricting attention to a suitable σ-algebra (the Lebesgue measurable sets), which is the object to be constructed next.

§4. 16 September 2026

Having seen that no measure on *all* subsets of $\mathbb{R}$ can extend length while remaining translation invariant, we now begin the actual construction of Lebesgue measure. The construction proceeds geometrically, starting from rectangles (objects simple enough that their "volume" is unambiguous) and builds up to general open, and eventually general measurable, sets. This lecture develops the purely geometric facts about rectangles and open sets that the construction will rely on.

Definition 4.1 (Rectangle)**.** A (closed) **rectangle** $R \subseteq \mathbb{R}^d$ is a product of d closed, bounded intervals:

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d], \qquad a_i \leq b_i. \tag{4.1}$$

The sides of R are parallel to the coordinate axes; rotated rectangles are not considered. The associated **open rectangle** is

$$R^\circ := (a_1, b_1) \times (a_2, b_2) \times \cdots \times (a_d, b_d). \tag{4.2}$$

Definition 4.2 (Volume)**.** The **volume** of the rectangle $R = \prod_{i=1}^d [a_i, b_i]$ is

$$|R| := \prod_{i=1}^{d} (b_i - a_i). \tag{4.3}$$

(The same formula, with the same value, defines the volume $|R^\circ|$ of the associated open rectangle.)

Definition 4.3 (Almost disjoint collection)**.** A finite (or countable) collection of rectangles $\{R_k\}$ is **almost disjoint** if the open rectangles R_k° are pairwise disjoint; that is, the rectangles may touch along their boundaries (share a face, edge, or corner), but their interiors do not overlap.

§4.i. Additivity of Volume over Almost Disjoint Partitions

Proposition 4.1

Suppose a rectangle R is the almost disjoint union of finitely many rectangles $R_1, \ldots, R_n$, i.e. $R = \bigcup_{k=1}^n R_k$ with $\{R_k\}$ almost disjoint. Then

$$|R| = \sum_{k=1}^{n} |R_k| \,. \tag{4.4}$$

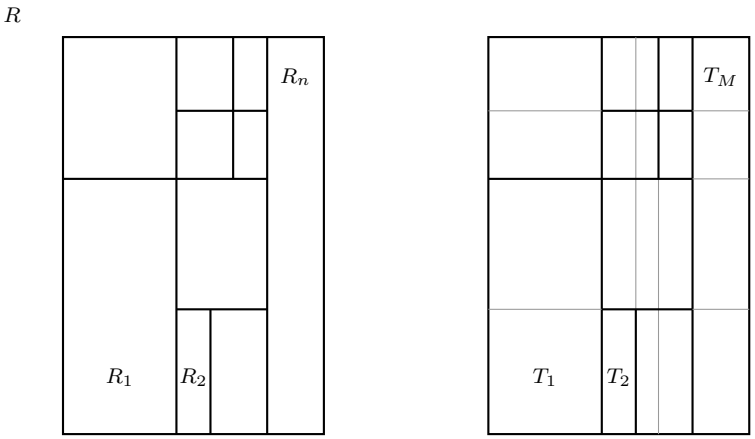
Proof. **Case 1:** Suppose first that each side of $R = \prod_{i=1}^d [a_i, b_i]$ has been partitioned into finitely many subintervals. Say the i-th side $[a_i, b_i]$ is cut into pieces of lengths $\ell_{i,1}, \ldots, \ell_{i,m_i}$

with $\sum_j \ell_{i,j} = b_i - a_i$, and the R_k are exactly the resulting grid cells (one for each choice of a subinterval in each coordinate). Then, using distributivity of multiplication over addition in each coordinate,

$$|R| = \prod_{i=1}^{d}(b_i - a_i) = \prod_{i=1}^{d}\left(\sum_{j=1}^{m_i} \ell_{i,j}\right) = \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \ell_{1,j_1} \cdots \ell_{d,j_d} = \sum_k |R_k|\,, \tag{4.5}$$

since each term in the expanded product is exactly the volume of one grid cell, and every grid cell arises exactly once.

Case 2: Given an arbitrary almost disjoint decomposition $R = \bigcup_{k=1}^{n} R_k$, extend every face (bounding hyperplane) of every R_k across the entirety of R, in every coordinate direction.



This produces a common refining grid on R: each original rectangle R_k is itself partitioned by this grid into finitely many grid cells (an almost disjoint union of sub-rectangles), and by Case 1 applied to R_k,

$$|R_k| = \sum_{\substack{\text{grid cells } T \\ T \subseteq R_k}} |T|\,. \tag{4.6}$$

On the other hand, since $\{R_k\}$ is almost disjoint and covers R, every grid cell T of the common refinement lies in exactly one R_k (up to boundary overlaps, which do not affect volume), so summing over k,

$$\sum_{k=1}^{n} |R_k| = \sum_{k=1}^{n} \sum_{T \subseteq R_k} |T| = \sum_{\text{all grid cells } T \text{ of } R} |T| = |R|\,, \tag{4.7}$$

the last equality again by Case 1, applied this time to R itself with the common refining grid. ■

§4.ii. Subadditivity of Volume over Coverings

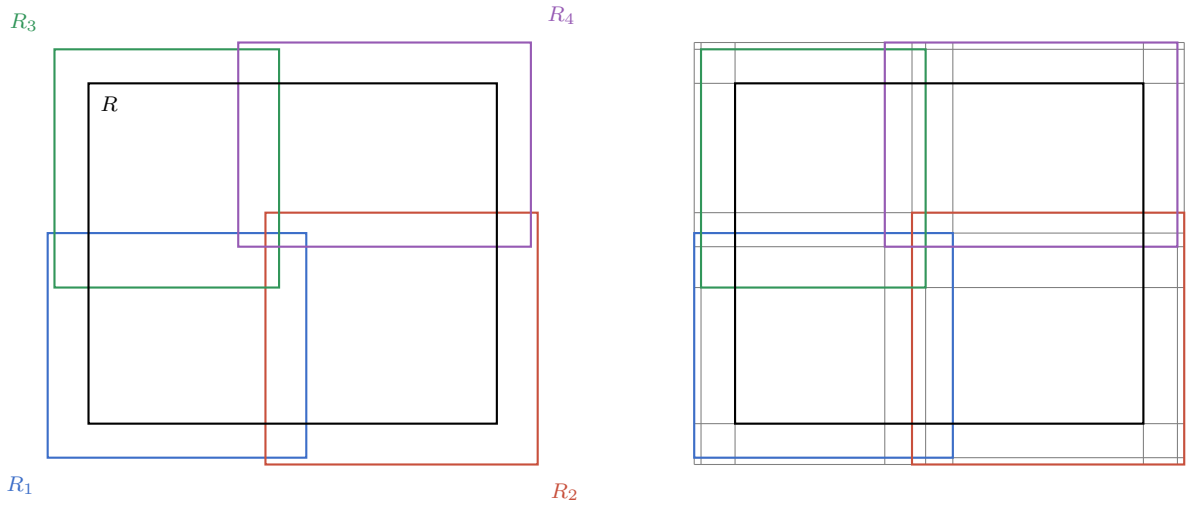
Proposition 4.2

Suppose a rectangle R is covered by finitely many rectangles $R_1, \ldots, R_n$ (i.e. $R \subseteq \bigcup_{k=1}^{n} R_k$), with no assumption that the R_k are disjoint or almost disjoint, and no assumption that

$R_k \subseteq R$. Then

$$|R| \leq \sum_{k=1}^{n} |R_k| \,. \tag{4.8}$$

Proof. As in the proof of Proposition 4.1, extend every face of every R_k and R to form a refining grid. This may create grid rectangles that are **NOT** in any of the R_i's or R.



This decomposes R into finitely many almost disjoint grid cells $T_1, \ldots, T_m$, so that $R = \bigcup_j T_j$ almost disjointly, and by Proposition 4.1,

$$|R| = \sum_{j=1}^{m} |T_j| \,. \tag{4.9}$$

Since $R \subseteq \bigcup_k R_k$, every grid cell T_j lies in at least one R_k; assign each T_j to one such R_k (breaking ties arbitrarily).

$$|R| = \sum_{j=1}^{m} |T_j| = \sum_{k=1}^{n} \left(\sum_{\substack{T_j \text{ assigned} \\ \text{to } R_k}} |T_j| \right) . \tag{4.10}$$

For a fixed k, the cells T_j assigned to R_k are almost disjoint and all contained in $R_k \cap R \subseteq R_k$. Note that they may not exhaust all of R_k. In fact, they are a subset of all the grid rectangles that are contained in R_k. Therefore,

$$\sum_{\substack{T_j \text{ assigned} \\ \text{to } R_k}} |T_j| \leq \sum_{\substack{T \text{ grid rectangle,} \\ T \subseteq R_k}} |T| = |R_k| \,, \tag{4.11}$$

where the last equality follows from Proposition 4.1. Now combining (4.10) and (4.11), we are done! ∎

Remark 4.1. Proposition 4.2 is the geometric heart of **countable subadditivity** for the outer measure to be constructed in the next lecture: it says that volume cannot be "cheated" by covering a rectangle with other rectangles of smaller total volume.

§4.iii. The Structure of Open Sets in $\mathbb{R}$

Proposition 4.3 (Structure of open subsets of $\mathbb{R}$)

Every open set $O \subseteq \mathbb{R}$ can be written as a countable union of pairwise disjoint open intervals. Moreover, this decomposition is unique: it is exactly the decomposition of O into its connected components.

Proof. For $x \in O$, let

$$I_x := (a_x, b_x)\,, \text{ where } a_x = \inf\{a : (a, x) \subseteq O\}\,, \text{ and } b_x = \sup\{b : (x, b) \subseteq O\}\,, \tag{4.12}$$

be the *largest* open interval contained in O that contains x. Since O is open, I_x is nonempty (it contains some small interval around x), and clearly $O = \bigcup_{x \in O} I_x$.

If $x, y \in O$ and $I_x \cap I_y \neq \varnothing$, then $I_x \cup I_y$ is itself an open interval (a union of two overlapping intervals) contained in O and containing both x and y; by maximality of I_x (as the *largest* such interval containing x), we must have

$$I_x \cup I_y \subseteq I_x \implies I_y \subseteq I_x. \tag{4.13}$$

Similarly, $I_x \subseteq I_y$, so $I_x = I_y$. Hence, the collection $\{I_x : x \in O\}$ consists of intervals that are pairwise either identical or disjoint; selecting one representative from each distinct interval gives a collection of pairwise disjoint open intervals $\{I_j\}_{j \in J}$ with $O = \bigsqcup_{j \in J} I_j$.

Each I_j is a nonempty open interval, hence contains a rational number q_j; since the I_j are pairwise disjoint, distinct intervals contain distinct rationals, giving an injection

$$\begin{aligned} J &\to \mathbb{Q} \\ j &\mapsto q_j. \end{aligned} \tag{4.14}$$

An injection into a countable set shows the index set J is countable.

Now about the uniqueness, suppose

$$O = \bigsqcup_k J_k, \tag{4.15}$$

is another decomposition into pairwise disjoint open intervals. Each J_k, being a connected subset of O, is contained in the maximal interval through any of its points, i.e. $J_k \subseteq I_j$ for a **unique** j.

Conversely, I_j is covered by those J_k's contained in it. But I_j is connected and the J_k's are open and pairwise disjoint, so I_j cannot be covered by two or more nonempty disjoint open pieces (else I_j would be disconnected).

Hence exactly one J_k satisfies $J_k \subseteq I_j$ with the covering forcing $J_k = I_j$. This gives a bijection between the two decompositions, proving uniqueness. ■

§4.iv. The Structure of Open Sets in $\mathbb{R}^d$, $d \geq 2$

Proposition 4.3 is special to $\mathbb{R}$: it relies essentially on the fact that connected open subsets of $\mathbb{R}$ are exactly the open intervals. In $\mathbb{R}^d$ for $d \geq 2$, connected open sets can have arbitrarily complicated shapes (e.g. an open disk), so an open set in $\mathbb{R}^d$ cannot in general be written as a disjoint union of open *rectangles*. The right substitute uses **closed rectangles**, allowing *almost* disjointness in place of disjointness.

Remark 4.2. In $d = 1$, consider the open set $(0, 1) \subseteq \mathbb{R}$. Then we can write $(0, 1)$ as a countable union of almost-disjoint closed intervals:

$$(0,1) = \bigcup_{n=2}^{\infty} \left[\frac{1}{n+1}, \frac{1}{n}\right] \cup \bigcup_{n=2}^{\infty} \left[1 - \frac{1}{n}, 1 - \frac{1}{n+1}\right]. \tag{4.16}$$

Definition 4.4 (Dyadic cubes). For $n \in \mathbb{Z}$, a **dyadic cube of generation** n is a set of the form

$$Q = \prod_{i=1}^{d} \left[\frac{k_i}{2^n}, \frac{k_i + 1}{2^n}\right], \qquad k_1, \ldots, k_d \in \mathbb{Z}. \tag{4.17}$$

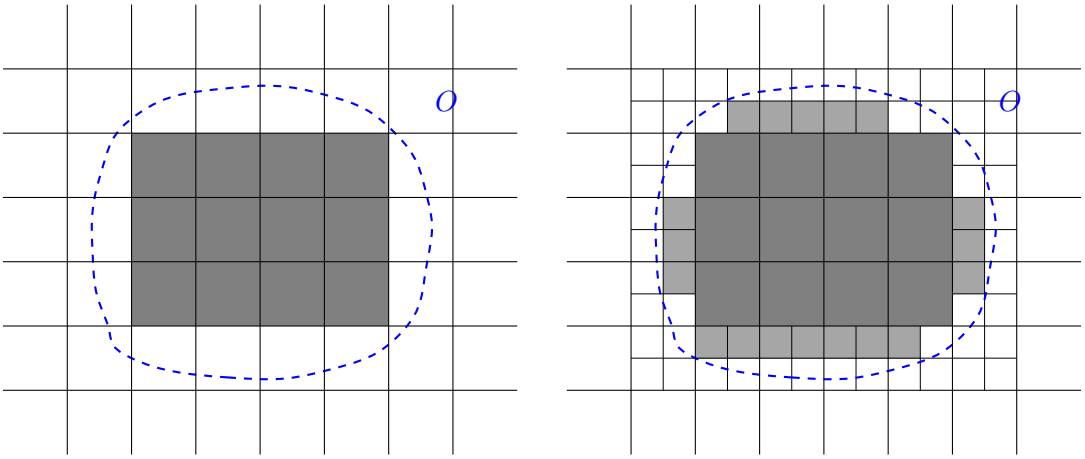
For fixed n, the dyadic cubes of generation n have disjoint interiors and their union is all of $\mathbb{R}^d$ (they form a grid of cubes of side length 2^{-n}). Every dyadic cube Q of generation n is contained in a unique dyadic cube $\widetilde{Q}$ of generation $n - 1$ (its **parent**, of side length $2^{-(n-1)}$), obtained by halving the grid spacing; conversely, each dyadic cube of generation $n - 1$ is the almost disjoint union of exactly 2^d dyadic cubes of generation n (its **children**).

Remark 4.3. Two dyadic cubes (of possibly different generations) are always either almost disjoint (disjoint interiors) or one is contained in the other. This nesting property is the key structural fact that makes the dyadic grid useful.

Proposition 4.4 (Structure of open subsets of $\mathbb{R}^d$)

Every open set $O \subseteq \mathbb{R}^d$ can be written as a countable, almost disjoint union of closed rectangles.

Proof. Call a dyadic cube Q **selected** if $Q \subseteq O$ but its parent $\widetilde{Q}$ is *not* contained in O (i.e. Q is a maximal dyadic cube contained in O, in the sense that enlarging it to its parent would exit O).



Define

$$\mathcal{Q} := \{Q : Q \text{ is a selected dyadic cube}\}, \qquad U := \bigcup_{Q \in \mathcal{Q}} Q. \tag{4.18}$$

We claim $O = U$. $U \subseteq O$ is immediate, since every selected Q satisfies $Q \subseteq O$ by definition.

Let $x \in O$. Since O is open, there is $\varepsilon > 0$ with the open ball $B(x, \varepsilon) \subseteq O$. Choose n large enough that a dyadic cube of generation n has diameter

$$2^{-n}\sqrt{d} < \varepsilon. \tag{4.19}$$

Then there is a generation-n dyadic cube Q_n containing x that satisfies

$$Q_n \subseteq B(x, \varepsilon) \subseteq O. \tag{4.20}$$

Since $Q_n \subseteq O$, that means Q_n has been selected at step n, or one of its previous generation parents has been selected at an earlier step. Therefore, $x \in U = \bigcup_{Q \in \mathcal{Q}} Q$.

Now we show that $\mathcal{Q}$ is an almost-disjoint collection. Suppose $Q, Q' \in \mathcal{Q}$ are distinct selected cubes with overlapping interiors. By the nesting property, one is contained in the other; say

$$Q' \subsetneq Q, \tag{4.21}$$

so that Q' is a proper descendant of Q, of strictly higher generation. Let Q'' be the immediate parent of Q'. Then

$$Q' \subseteq Q'' \subseteq Q. \tag{4.22}$$

Since $Q \subseteq O$ (as Q is selected), also $Q'' \subseteq O$. But this contradicts Q' being selected, which requires the immediate parent of Q' to *not* be contained in O. Hence no such containment can occur, and the interiors of distinct selected cubes are disjoint.

Finally, we'll now show that $\mathcal{Q}$ is a countable collection. For each fixed generation $n \in \mathbb{Z}$, the dyadic cubes of generation n are indexed by $\mathbb{Z}^d$, a countable set.

$$\mathcal{Q} \subseteq \bigcup_{n \in \mathbb{Z}} \{\text{generation-}n \text{ dyadic cubes}\}, \tag{4.23}$$

which is a countable union of countable sets, hence countable. ■

Remark 4.4 (Comparison with the one-dimensional case)**.** For $d = 1$, the construction in Proposition 4.4 produces a decomposition into closed dyadic *intervals*, which is generally coarser and less canonical than the connected-components decomposition of Proposition 4.3 (dyadic intervals need not align with the true connected components of O, and distinct choices of the underlying grid (e.g. shifting or rescaling it) give different, equally valid, decompositions). This is consistent with the failure of *uniqueness* in higher dimensions: Proposition 4.4 asserts existence of an almost disjoint decomposition into closed rectangle, not that it is the unique or canonical one.

Remark 4.5. The decomposition of Proposition 4.4 cannot in general be made into a genuinely disjoint union of *closed* cubes (closed cubes covering a connected open set must share boundary faces where they meet), nor into a disjoint union of *open* rectangles when $d \geq 2$ (e.g. an open disk is not a union of finitely or countably many disjoint open rectangles, since removing the boundary of any inscribed rectangle from a disk leaves a connected leftover region that cannot itself be a union of open rectangles disjoint from the first without gaps). Allowing cubes to touch along boundaries ("almost" disjoint) is what makes the construction possible in every dimension.

§5. 17 September 2026

We now begin the construction proper. The strategy has two stages: first, define a set function m^*, the **exterior measure** (or **outer measure**), on *every* subset of $\mathbb{R}^d$ by covering from outside with cubes; second, single out a sub-collection of "good" sets (we'll call them the **Lebesgue measurable sets**) on which m^* will turn out to be countably additive, i.e. a genuine measure in the sense of Section 3. Theorem 3.11 tells us this sub-collection cannot be all of $\mathcal{P}(\mathbb{R}^d)$.

§5.i. Exterior Measure

Definition 5.1 (Exterior measure)**.** For a set $E \subseteq \mathbb{R}^d$, the **exterior measure** of E is

$$m^*(E) := \inf\left\{ \sum_{j=1}^{\infty} |Q_j| \;:\; E \subseteq \bigcup_{j=1}^{\infty} Q_j, \quad Q_j \text{ closed cubes} \right\} \in [0,\infty]. \tag{5.1}$$

That is, $m^*(E)$ is the infimum, over all countable covers of E by closed cubes (not required to be disjoint or almost disjoint), of the total volume of the cubes used.

Remark 5.1. Every set $E \subseteq \mathbb{R}^d$ admits at least one such cover (e.g. by cubes tiling all of $\mathbb{R}^d$), so $m^*(E)$ is always a well-defined element of $[0,\infty]$.

Remark 5.2 (Why cover from outside)**.** One might instead try to define an "interior" measure by taking the *supremum* of the total volume of cubes contained *inside* E. This alternative fails badly. For example, a set with empty interior but which is, intuitively, "large" (such as the set of irrational numbers in $[0,1]$, or more strikingly a dense set with no interior points at all) would have interior measure 0, since no cube of positive volume can fit inside it. Covering from *outside* does not suffer from this defect, which is why the construction proceeds this way.

Remark 5.3 (Cubes versus rectangles)**.** The definition uses closed cubes rather than closed rectangles purely for concention. Using rectangles in place of cubes gives an identical set function (the two notions of covering are interchangeable, since every rectangle can be covered by finitely many cubes, and vice versa). Open cubes could equally be used in place of closed cubes, again without changing the value of m^*.

Example 5.1

$m^*(\{x\}) = 0$ for any single point $x \in \mathbb{R}^d$: cover $\{x\}$ by a closed cube of side length ε centered at x, of volume ε^d. Since $\varepsilon > 0$ is arbitrary, the infimum over such covers is 0.

Remark 5.4. Since $m^*(E) = \inf_{E \subseteq \bigcup_{j=1}^{\infty} Q_j} \sum_{j=1}^{\infty} |Q_j|$, by the definition of infimum, for every $\varepsilon > 0$, we can find a cover using closed cubes $E \subseteq \bigcup_{j=1}^{\infty} Q_j$ such that

$$\sum_{j=1}^{\infty} |Q_j| \leq m^*(E) + \varepsilon. \tag{5.2}$$

This will turn out to be helpful while proving some identities.

§5.ii. Basic Properties of Exterior Measure

We record some basic properties of m^*. These would guide us toward the properties we ultimately want a measure to have.

Proposition 5.1 (Monotonicity)

If $E \subseteq F \subseteq \mathbb{R}^d$, then $m^*(E) \leq m^*(F)$.

Proof. Every countable cover of F by closed cubes is also a cover of E (since $E \subseteq F$), so the collection of covers admissible for computing $m^*(F)$ is a subset of those admissible for $m^*(E)$. An infimum taken over a smaller collection of covers cannot be less than one taken over a larger collection, i.e. $m^*(E) \leq m^*(F)$. ■

Proposition 5.2 (Countable subadditivity)

For any countable collection of sets $E_1, E_2, \ldots \subseteq \mathbb{R}^d$,

$$m^*\left(\bigcup_{j=1}^{\infty} E_j\right) \leq \sum_{j=1}^{\infty} m^*(E_j). \tag{5.3}$$

Proof. If the right side is infinite there is nothing to prove, so assume it is finite. Fix $\varepsilon > 0$. For each j, by definition of $m^*(E_j)$ as an infimum, there is a countable cover of E_j by closed cubes $\{Q_{j,i}\}_{i \geq 1}$ with

$$\sum_{i=1}^{\infty} |Q_{j,i}| \leq m^*(E_j) + \frac{\varepsilon}{2^j}. \tag{5.4}$$

The combined collection $\{Q_{j,i}\}_{j,i \geq 1}$ is a countable cover of $\bigcup_j E_j$ by closed cubes, so by definition of $m^*\left(\bigcup_j E_j\right)$ as an infimum,

$$m^*\left(\bigcup_{j=1}^{\infty} E_j\right) \leq \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} |Q_{j,i}| \leq \sum_{j=1}^{\infty}\left(m^*(E_j) + \frac{\varepsilon}{2^j}\right) = \sum_{j=1}^{\infty} m^*(E_j) + \varepsilon. \tag{5.5}$$

Since $\varepsilon > 0$ was arbitrary, the claimed inequality follows. ■

Corollary 5.3

If E is countable, then $m^*(E) = 0$.

Proof. Suppose $E = \{x_1, x_2, x_3, \ldots\}$. Then

$$E = \bigcup_{i=1}^{\infty} \{x_i\} \implies 0 \leq m^*(E) \leq \sum_{i=1}^{\infty} m^*(\{x_i\}) = 0. \tag{5.6}$$

■

Proposition 5.4 (Exterior measure of a cube equals its volume)

If Q is a closed cube, then $m^*(Q) = |Q|$.

Proof. The single-cube cover $\{Q\}$ is admissible, giving $m^*(Q) \leq |Q|$ directly from the definition.

Now we shall prove that $m^*(Q) \geq |Q|$. Let $\{Q_j\}_{j\geq 1}$ be any countable cover of Q by closed cubes; we must show $\sum_{j=1}^{\infty} |Q_j| \geq |Q|$.

Fix $\varepsilon > 0$, and for each j let $\widetilde{Q}_j^\circ$ be the open cube with the same center as Q_j, dilated by a factor $(1+\varepsilon)$, i.e.

$$\widetilde{Q}_j^\circ = (1+\varepsilon)\, Q_j^\circ, \tag{5.7}$$

so that $Q_j \subset \widetilde{Q}_j^\circ$ and

$$\left|\widetilde{Q}_j^\circ\right| = \left|\widetilde{Q}_j\right| = (1+\varepsilon)^d \left|Q_j\right|. \tag{5.8}$$

Then $\{\widetilde{Q}_j^\circ\}$ is a cover of Q by open sets, and since Q is compact, finitely many of them already cover Q, say $Q \subseteq \bigcup_{j=1}^{n} \widetilde{Q}_j^\circ$. By Proposition 4.2 (subadditivity of volume for coverings of a rectangle),

$$|Q| \;\leq\; \sum_{j=1}^{n} \left|\widetilde{Q}_j^\circ\right| = (1+\varepsilon)^d \sum_{j=1}^{n} |Q_j| \;\leq\; (1+\varepsilon)^d \sum_{j=1}^{\infty} |Q_j|. \tag{5.9}$$

Since $\varepsilon > 0$ is arbitrary, letting $\varepsilon \to 0^+$ gives $|Q| \leq \sum_{j=1}^{\infty} |Q_j|$. As this holds for every admissible cover $\{Q_j\}$, taking the infimum over all such covers gives $|Q| \leq m^*(Q)$. ■

Proposition 5.5

If Q° is an open cube, then $m^*(Q^\circ) = |Q^\circ|$.

Proof. Let Q be the closure of Q°. Then $Q^\circ \subseteq Q$ and monotonicity (Proposition 5.1) gives us

$$m^*(Q^\circ) \leq m^*(Q) = |Q| = |Q^\circ|. \tag{5.10}$$

Now note that $(1-\varepsilon)\, Q \subseteq Q^\circ$, and hence, monotonicity will give us

$$m^*(Q^\circ) \geq m^*((1-\varepsilon)\, Q) = |(1-\varepsilon)\, Q| = (1-\varepsilon)^d \left|Q\right| = (1-\varepsilon)^d \left|Q^\circ\right|. \tag{5.11}$$

Letting $\varepsilon \to 0^+$ forces $m^*(Q^\circ) \geq |Q^\circ|$. ■

The same argument (covering by cubes and using the volume-subadditivity Proposition 4.2) shows more generally that $m^*(R) = |R|$ for any rectangle R.

Proposition 5.6

For any rectangle R, $m^*(R) = |R|$.

Proof. Let $\{Q_j\}_{j\geq 1}$ be any countable cover of R by closed cubes. Fix $\varepsilon > 0$, and for each j let

$$\widetilde{Q}_j^\circ = (1+\varepsilon)\, Q_j^\circ, \tag{5.12}$$

so that $R \subseteq \bigcup_{j=1}^\infty Q_j \subseteq \bigcup_{j=1}^\infty \widetilde{Q}_j^\circ$. By the compactness of R, we have $R \subseteq \bigcup_{j=1}^n \widetilde{Q}_j^\circ$. Using Proposition 4.2, we get

$$|R| \;\leq\; \sum_{j=1}^n \left|\widetilde{Q}_j^\circ\right| = (1+\varepsilon)^d \sum_{j=1}^n |Q_j| \;\leq\; (1+\varepsilon)^d \sum_{j=1}^\infty |Q_j|\,. \tag{5.13}$$

Since $\varepsilon > 0$ is arbitrary, letting $\varepsilon \to 0^+$ gives $|Q| \leq \sum_{j=1}^\infty |Q_j|$. As this holds for every admissible cover $\{Q_j\}$, taking the infimum over all such covers gives $|R| \leq m^*(R)$.

Now we'll show that $m^*(R) \leq |R|$. Consider a grid in $\mathbb{R}^d$ formed by cubes of size $\frac{1}{k}$.

$$\mathcal{Q} := \{Q : Q \text{ is a cube contained in } R\}\,, \tag{5.14}$$

$$\mathcal{Q}' := \{Q : Q \text{ is a cube that intersect both } R \text{ and } R^c\}\,. \tag{5.15}$$

Then we have

$$\bigcup_{Q\in\mathcal{Q}} Q \subseteq R \subseteq \bigcup_{Q\in\mathcal{Q}\cup\mathcal{Q}'} Q. \tag{5.16}$$

R is made out of $\bigcup_{Q\in\mathcal{Q}} Q$ (which itself is a collection of almost-disjoint rectangles) and some more almost-disjoint rectangles. Therefore,

$$|R| = \sum_{Q\in\mathcal{Q}} |Q| + \text{ possibly some more } \geq \sum_{Q\in\mathcal{Q}} |Q|\,. \tag{5.17}$$

Claim 5.1: The size of $\mathcal{Q}'$ is at most Ck^{d-1}, for some constant C (not depending on k).

Proof. The cubes in $\mathcal{Q}'$ intersect both R and R^c, i.e. they cover the $(d-1)$-dimensional faces of R. There are exactly $2d$ of them.

Let a be the largest edge-length of R. To cover the edge of length a, we would need around $\frac{a}{k^{-1}} = ak$ cubes. But this would be the case if the cubes fit perfectly into the edge. There may be a little space leftover, so the number of cubes we need is no more than $ak+2$ (adding one for each endpoint of the edge). Therefore, to cover a $(d-1)$-dimensional face

of R, the number of cubes needed is at most

$$(ak+2)^{d-1} = (a+1)^{d-1}\, k^{d-1}. \tag{5.18}$$

There are $2d$ faces, so the number of cubes needed to cover the faces is at most

$$2d \cdot (a+1)^{d-1}\, k^{d-1}. \tag{5.19}$$

□

Then we have $R \subseteq \bigcup_{Q\in\mathcal{Q}\cup\mathcal{Q}'} Q$, and each cube has volume $\frac{1}{k^d}$. Therefore,

$$\sum_{Q\in\mathcal{Q}\cup\mathcal{Q}'} |Q| = \sum_{Q\in\mathcal{Q}} |Q| + \sum_{Q\in\mathcal{Q}'} |Q| \le |R| + Ck^{d-1}\cdot\frac{1}{k^d} = |R| + \frac{C}{k}. \tag{5.20}$$

Taking $k\to\infty$, we get $m^*(R) \le |R|$. ■

Proposition 5.7

For any $E \subseteq \mathbb{R}^d$,

$$m^*(E) = \inf\{m^*(O) : E \subseteq O,\ O \text{ open}\}. \tag{5.21}$$

Proof. For any open $O \supseteq E$, monotonicity (Proposition 5.1) gives $m^*(E) \le m^*(O)$; taking the infimum over such O preserves the inequality:

$$m^*(E) \le \inf_{\substack{E\subseteq O,\\ O \text{ open}}} m^*(O). \tag{5.22}$$

Now, for the other inequality: fix $\varepsilon > 0$. By definition of $m^*(E)$, there is a countable cover of E by closed cubes $\{Q_j\}$ with

$$\sum_{j=1}^{\infty} |Q_j| \le m^*(E) + \varepsilon. \tag{5.23}$$

For each j, let

$$\widetilde{Q}_j^\circ = (1+\varepsilon)\, Q_j^\circ, \qquad \text{and} \qquad O := \bigcup_j \widetilde{Q}_j^\circ, \tag{5.24}$$

which is an open set containing E (since $E \subseteq \bigcup_j Q_j \subseteq \bigcup_j \widetilde{Q}_j^\circ$). By countable subadditivity (Proposition 5.2) and Proposition 5.4,

$$m^*(O) \le \sum_{j=1}^{\infty} m^*(\widetilde{Q}_j^\circ) = \sum_{j=1}^{\infty} \left|\widetilde{Q}_j^\circ\right| \ \le\ (1+\varepsilon)^d \sum_{j=1}^{\infty} |Q_j| \le (1+\varepsilon)^d\, m^*(E) + \varepsilon. \tag{5.25}$$

Taking infimum on the LHS gives

$$\inf_{\substack{E\subseteq O,\\ O \text{ open}}} m^*(O) \le (1+\varepsilon)^d\, m^*(E) + \varepsilon, \tag{5.26}$$

for every $\varepsilon > 0$. Now taking $\varepsilon \to 0^+$ gives us the desired bound. ■

Definition 5.2. The diameter of a set $X \subseteq \mathbb{R}^d$ is

$$\operatorname{diam}(X) = \sup\{|x - y| : x, y \in X\}. \tag{5.27}$$

The distance between two sets $X, Y \subseteq \mathbb{R}^d$ is

$$d(X, Y) := \inf\{|x - y| : x \in X, y \in Y\}. \tag{5.28}$$

Proposition 5.8 (Additivity for positively separated sets)

If $d(E_1, E_2) > 0$, then

$$m^*(E_1 \cup E_2) = m^*(E_1) + m^*(E_2). \tag{5.29}$$

Proof. Applying subadditivity (Proposition 5.2) to the two-set collection $\{E_1, E_2\}$, we get

$$m^*(E_1 \cup E_2) \leq m^*(E_1) + m^*(E_2). \tag{5.30}$$

For the other inequality, we may assume $m^*(E_1 \cup E_2) < \infty$ (else there is nothing to prove). Fix $\varepsilon > 0$ and let $\{Q_j\}$ be a cover of $E_1 \cup E_2$ by closed cubes with

$$\sum_{j=1}^{\infty} |Q_j| \leq m^*(E_1 \cup E_2) + \varepsilon. \tag{5.31}$$

By subdividing each Q_j into finitely many smaller closed cubes if necessary (which does not change the total volume, by Proposition 4.1), we may further assume every cube in the cover has diameter less than δ, with

$$0 < \delta < d(E_1, E_2). \tag{5.32}$$

Then no single cube Q_j can intersect both E_1 and E_2; indeed if Q_i intersects both E_1 and E_2 at points e_1 and e_2, respectively, then

$$|e_1 - e_2| \leq \operatorname{diam}(Q_j) \leq \delta < d(E_1, E_2) \leq |e_1 - e_2|, \tag{5.33}$$

which is not possible. Now, we partition the cover into $\{Q_j : Q_j \cap E_1 \neq \varnothing\}$, which covers E_1, and $\{Q_j : Q_j \cap E_2 \neq \varnothing\}$, which covers E_2 (every cube meets at most one of the two sets, so this is a genuine partition of the index set). Then

$$m^*(E_1) \leq \sum_{Q_j \cap E_1 \neq \varnothing} |Q_j| \quad \text{and} \quad m^*(E_2) \leq \sum_{Q_j \cap E_2 \neq \varnothing} |Q_j|. \tag{5.34}$$

Adding these, we get

$$m^*(E_1) + m^*(E_2) \leq \sum_{Q_j \cap E_1 \neq \varnothing} |Q_j| + \sum_{Q_j \cap E_2 \neq \varnothing} |Q_j| = \sum_{j=1}^{\infty} |Q_j| \leq m^*(E_1 \cup E_2) + \varepsilon. \tag{5.35}$$

Since $\varepsilon > 0$ is arbitrary, $m^*(E_1) + m^*(E_2) \leq m^*(E_1 \cup E_2)$. ■

Corollary 5.9

If $E_1, E_2, \ldots, E_n \subseteq \mathbb{R}^d$ with $d(E_i, E_j) > 0$ for every $i \neq j$, then

$$m^*\left(\bigcup_{i=1}^{n} E_n\right) = \sum_{i=1}^{n} m^*(E_i). \tag{5.36}$$

Proposition 5.10 (Exterior measure of an almost disjoint union of cubes)

If $E = \bigcup_{j=1}^{\infty} Q_j$ is a countable, almost disjoint union of closed cubes, then

$$m^*(E) = \sum_{j=1}^{\infty} |Q_j|. \tag{5.37}$$

Proof. $m^*(E) \leq \sum_{j=1}^{\infty} |Q_j|$ directly follows from the definition, since $\bigcup_{j=1}^{\infty} Q_j$ is countable collection of closed cubes that cover E. For the other direction, fix $N \geq 1$ and $\varepsilon > 0$, and let

$$\widetilde{Q}_j = (1 - \varepsilon) Q_j. \tag{5.38}$$

Since $Q_1, \ldots, Q_N$ are almost disjoint (disjoint interiors) and there are only finitely many of them, shrinking creates a positive gap between each pair, i.e. for $i \neq j$,

$$d(Q_i', Q_j') > 0. \tag{5.39}$$

Corollary 5.9 together with Proposition 5.4 gives

$$m^*\left(\bigcup_{j=1}^{N} \widetilde{Q}_j\right) = \sum_{j=1}^{N} m^*(\widetilde{Q}_j) = \sum_{j=1}^{N} \left|\widetilde{Q}_j\right| = (1-\varepsilon)^d \sum_{j=1}^{N} |Q_j|. \tag{5.40}$$

Since $\bigcup_{j=1}^{N} Q_j' \subseteq E$, monotonicity (Proposition 5.1) gives

$$m^*(E) \geq m^*\left(\bigcup_{j=1}^{N} \widetilde{Q}_j\right) = (1-\varepsilon)^d \sum_{j=1}^{N} |Q_j|. \tag{5.41}$$

Letting $\varepsilon \to 0^+$ (with N fixed), so $m^*(E) \geq \sum_{j=1}^{N} |Q_j|$. Since this holds for every N, letting $N \to \infty$ gives $m^*(E) \geq \sum_{j=1}^{\infty} |Q_j|$. ■

Remark 5.5. Proposition 5.10 combined with Proposition 4.4 (every open set is a countable almost disjoint union of dyadic cubes), gives a concrete way of computing $m^*(O)$ for an open set O. That is, decompose O into almost disjoint dyadic cubes and sum their volumes.

§5.iii. Lebesgue Measurable Sets

We now single out the sets on which m^* will be shown (in the next lecture) to be countably additive.

Definition 5.3 (Lebesgue measurable set)**.** A set $E \subseteq \mathbb{R}^d$ is **Lebesgue measurable** (or simply **measurable**, in this context) if, for every $\varepsilon > 0$, there exists an open set $O \supseteq E$ such that

$$m^*(O - E) \leq \varepsilon. \tag{5.42}$$

Remark 5.6. Informally, E is measurable if it can be approximated from outside by open sets, in the sense that the "error" $O - E$ can be made to have arbitrarily small exterior measure. This definition makes no direct reference to a σ-algebra. We will show below that the collection of measurable sets does in fact form a σ-algebra.

Remark 5.7. Proposition 5.7 says that for every $\varepsilon > 0$, we can find an open set $O \supseteq E$ such that

$$m^*(O) \leq m^*(E) + \varepsilon. \tag{5.43}$$

Note that this is quite different from (5.42), since it's **NOT** always true that $m^*(O - E) = m^*(O) - m^*(E)$.

We now establish a few basic properties of Lebesgue measurable sets, culminating in the fact that they form a σ-algebra.

Proposition 5.11 (Open sets are measurable)

Every open set $O \subseteq \mathbb{R}^d$ is measurable.

Proof. Take the open set to be O itself. Then $m^*(O - O) = m^*(\varnothing) = 0 < \varepsilon$ for every $\varepsilon > 0$. ∎

Proposition 5.12 (Sets of exterior measure zero are measurable)

If $m^*(E) = 0$, then E is measurable. (In particular, $m^*(E) = 0$ forces E to be measurable with m^* restricting to a genuine value 0 on E; such sets are called **null sets**.)

Proof. Fix $\varepsilon > 0$. By Proposition 5.7 applied to E, since $m^*(E) = 0$, there is an open set $O \supseteq E$ with $m^*(O) \leq \varepsilon$. Since $O - E \subseteq O$, monotonicity gives

$$m^*(O - E) \leq m^*(O) \leq \varepsilon. \tag{5.44}$$

∎

Proposition 5.13 (Countable unions of measurable sets are measurable)

If $E_1, E_2, \ldots$ are measurable, then $E := \bigcup_{n=1}^{\infty} E_n$ is measurable.

Proof. Fix $\varepsilon > 0$. For each n, since E_n is measurable, choose an open set $O_n \supseteq E_n$ with

$$m^*(O_n - E_n) \leq \frac{\varepsilon}{2^n}. \tag{5.45}$$

Let $O := \bigcup_{n=1}^{\infty} O_n$, which is an open set containing E. Since

$$O - E \subseteq \bigcup_{n=1}^{\infty} (O_n - E_n) \tag{5.46}$$

(a point in O not in E lies in some O_n, and since it is not in $E \supseteq E_n$, it is not in E_n either, so it lies in $O_n - E_n$), countable subadditivity (Proposition 5.2) gives

$$m^*(O - E) \leq \sum_{n=1}^{\infty} m^*(O_n - E_n) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon. \tag{5.47}$$

■

Proposition 5.14 (Closed sets are measurable)

Every closed set $F \subseteq \mathbb{R}^d$ is measurable.

Proof. **Case 1: Compact Sets.** Let K be compact (closed and bounded), we'll show that it's measurable. Fix $\varepsilon > 0$. By Proposition 5.7, there is an open set $O \supseteq K$ with

$$m^*(O) \leq m^*(K) + \varepsilon. \tag{5.48}$$

Since K is bounded, $m^*(K) < \infty$, so this is a genuine (finite) bound. The set $O - K$ is open (as K is closed), so by Proposition 4.4 it may be written as a countable almost disjoint union of closed cubes,

$$O - K = \bigcup_{n=1}^{\infty} Q_n. \tag{5.49}$$

For each fixed N, the finite union $S_N := \bigcup_{n=1}^{N} Q_n$ is compact (a finite union of closed bounded cubes), disjoint from K (since each $Q_n \subseteq O - K$), and two disjoint compact sets have positive distance; hence $d(K, S_N) > 0$ (see the following claim).

Claim 5.2: If $X, Y \subseteq \mathbb{R}^d$ are disjoint sets with X being compact and Y closed, then $d(X, Y) > 0$.

Proof. Assume for the sake of contradiction that there exists two sequences $x_n \in X$ and $y_n \in Y$ such that $|x_n - y_n| \to 0$. X is compact, so there exists a convergent subsequence

$x_{n_i} \to x \in X$. Then

$$|x - y_{n_i}| \leq |x - x_{n_i}| + |x_{n_i} - y_{n_i}|. \tag{5.50}$$

Both terms of RHS converges to 0. Therefore $|x - y_{n_i}| \to 0$. As Y is closed, it contains all the limit points, so $x \in Y$. But we are given that X and Y are disjoint. Contradiction! □

By Proposition 5.8 and Proposition 5.10,

$$m^*\left(K \cup \bigcup_{n=1}^{N} Q_n\right) = m^*(K) + m^*\left(\bigcup_{n=1}^{N} Q_n\right) = m^*(K) + \sum_{n=1}^{N} |Q_n|. \tag{5.51}$$

Since $K \cup \bigcup_{n=1}^{N} Q_n \subseteq O$, monotonicity gives

$$m^*(K) + \sum_{n=1}^{N} |Q_n| \leq m^*(O) \;\leq\; m^*(K) + \varepsilon \implies \sum_{n=1}^{N} |Q_n| \leq \varepsilon, \tag{5.52}$$

for every N, and letting $N \to \infty$,

$$\sum_{n=1}^{\infty} |Q_n| \leq \varepsilon. \tag{5.53}$$

By Proposition 5.10 again,

$$m^*(O - K) = m^*\left(\bigcup_{n=1}^{\infty} Q_n\right) = \sum_{n=1}^{\infty} |Q_n| \;\leq\; \varepsilon. \tag{5.54}$$

Since $\varepsilon > 0$ was arbitrary, K is measurable.

Case 2: General closed sets. Let F be closed (not necessarily bounded). Write

$$F = \bigcup_{n=1}^{\infty} \left(F \cap \overline{B(\mathbf{0}, n)}\right), \tag{5.55}$$

a countable union of compact sets (each $F \cap \overline{B(\mathbf{0}, n)}$ is closed and bounded). By Case 1, each is measurable, and by Proposition 5.13, their countable union F is measurable. ■

Proposition 5.15 (Complements of measurable sets are measurable)

If $E \subseteq \mathbb{R}^d$ is measurable, then $E^c = \mathbb{R}^d - E$ is measurable.

Proof. For each $n \geq 1$, since E is measurable, choose an open set $O_n \supseteq E$ with

$$m^*(O_n - E) \leq \frac{1}{n}. \tag{5.56}$$

Each O_n^c is closed, hence measurable by Proposition 5.14. Moreover, $O_n^c \subseteq E^c$, since $E \subseteq O_n$. Let

$$S := \bigcup_{n=1}^{\infty} O_n^c, \tag{5.57}$$

a countable union of measurable sets, hence measurable by Proposition 5.13. Since each $O_n^c \subseteq E^c$, also $S \subseteq E^c$. We claim $m^*(E^c - S) = 0$. Indeed,

$$E^c - S = E^c \cap S^c = E^c \cap \bigcap_{n=1}^{\infty} O_n = \bigcap_{n=1}^{\infty} (E^c \cap O_n) \subseteq E^c \cap O_n = O_n - E, \tag{5.58}$$

for every n. so

$$m^*(E^c - S) \leq m^*(O_n - E) \leq \frac{1}{n}, \tag{5.59}$$

for every n, forcing $m^*(E^c - S) = 0$. By Proposition 5.12, $E^c - S$ is measurable. Finally,

$$E^c = S \cup (E^c - S) \tag{5.60}$$

is a union of two measurable sets, hence measurable by Proposition 5.13. ■

Theorem 5.16 (The Lebesgue σ-algebra)

The collection $\mathcal{L}$ of Lebesgue measurable subsets of $\mathbb{R}^d$ is a σ-algebra.

Proof. Proposition 5.13 gives closure of $\mathcal{L}$ under countable unions, and Proposition 5.15 gives closure under complements. Furthermore, $\varnothing \in \mathcal{L}$ as it's a null-set. These are exactly the defining properties of a σ-algebra (Definition 1.1). ■

Remark 5.8. $\mathcal{L}$ contains every open set and hence contains the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. So in general

$$\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{L} \subsetneq \mathcal{P}(\mathbb{R}^d), \tag{5.61}$$

with the last containment strict by Vitali's theorem (Theorem 3.11).

Definition 5.4 (Lebesgue measure)**.** The **Lebesgue measure** of a set $E \in \mathcal{L}$ is defined as

$$m(E) := m^*(E), \tag{5.62}$$

i.e. m is the restriction of the exterior measure m^* to the σ-algebra $\mathcal{L}$ of measurable sets.

It remains to verify that m is in fact countably additive on $\mathcal{L}$, i.e. that $(\mathbb{R}^d, \mathcal{L}, m)$ is genuinely a measure space in the sense of Definition 1.3.

§6. 18 September 2026

§6.i. More properties of Lebesgue measure

The first order of business is to show that the set function m just defined is in fact a genuine *measure*: that it is countably additive on disjoint measurable sets.

Lemma 6.1

If E is measurable, then for every $\varepsilon > 0$, there exists a closed set $F \subseteq E$ such that $m^*(E - F) \leq \varepsilon$.

Proof. E^c is measurable, so there is an open set $O \supseteq E^c$ with $m(O - E^c) \leq \varepsilon$. Let $F = O^c$; then F is closed, and $F = O^c \subseteq (E^c)^c = E$, while

$$E - F = E \cap O = O - E^c, \tag{6.1}$$

so $m(E - F) = m(O - E^c) \leq \varepsilon$. ∎

Theorem 6.2 (Countable additivity)

If $E_1, E_2, E_3, \ldots$ is a countable collection of pairwise disjoint measurable sets, and $E = \bigcup_{j=1}^{\infty} E_j$, then E is measurable and

$$m(E) = \sum_{j=1}^{\infty} m(E_j). \tag{6.2}$$

Proof. Measurability of $E = \bigcup_j E_j$ is immediate, since $\mathcal{L}$ is already known to be a σ-algebra (closed under countable unions). Also, since the exterior measure is already known to be countably *sub*additive,

$$m(E) \leq \sum_{j=1}^{\infty} m(E_j). \tag{6.3}$$

So we'll now prove the other inequality. First, assume each E_j is bounded. Fix $\varepsilon > 0$. Since each E_j is measurable, there is a closed set $F_j \subseteq E_j$ with

$$m^*(E_j - F_j) \leq \frac{\varepsilon}{2^j}. \tag{6.4}$$

Since $E_j - F_j$ and F_j are disjoint with union E_j, using subadditivity,

$$m^*(E_j) \leq m^*(F_j) + m^*(E_j - F_j) \leq m^*(F_j) + \frac{\varepsilon}{2^j} \implies m(F_j) \geq m(E_j) - \frac{\varepsilon}{2^j}. \tag{6.5}$$

Now fix N and consider $F_1, \ldots, F_N$. Since $E_1, \ldots, E_N$ are pairwise disjoint, so are $F_1, \ldots, F_N$. Moreover, each F_j is closed and bounded (as $F_j \subseteq E_j$ and E_j is bounded), hence compact.

Applying Corollary 5.9 and Claim 5.2,

$$m^*\left(\bigcup_{j=1}^{N} F_j\right) = \sum_{j=1}^{N} m(F_j). \tag{6.6}$$

Since $\bigcup_{j=1}^{N} F_j \subseteq E$, monotonicity of outer measure gives

$$m(E) \;\geq\; m^*\left(\bigcup_{j=1}^{N} F_j\right) = \sum_{j=1}^{N} m(F_j) \;\geq\; \sum_{j=1}^{N} \left(m(E_j) - \frac{\varepsilon}{2^j}\right) \;\geq\; \sum_{j=1}^{N} m(E_j) \;-\; \varepsilon. \tag{6.7}$$

Letting $N \to \infty$ and then $\varepsilon \to 0$ (which was arbitrary) yields

$$m(E) \;\geq\; \sum_{j=1}^{\infty} m(E_j), \tag{6.8}$$

as desired. Now we move to the general scenario of when E_j's need not be bounded. Write $\mathbb{R}^d$ as an increasing union of closed balls: let B_k be the closed ball of radius k centered at the origin, and set

$$S_1 = B_1, \qquad S_k = B_k - B_{k-1} \quad (k \geq 2), \tag{6.9}$$

so that the S_k are pairwise disjoint and $\mathbb{R}^d = \bigcup_{k=1}^{\infty} S_k$. For each j, decompose E_j according to this partition:

$$E_{j,k} = E_j \cap S_k, \qquad E_j = \bigcup_{k=1}^{\infty} E_{j,k}, \tag{6.10}$$

a disjoint union of *bounded* measurable sets. By applying the bounded case to the sets $\{E_{j,k}\}_k$ (disjoint and bounded, for fixed j),

$$m(E_j) = \sum_{k=1}^{\infty} m(E_{j,k}). \tag{6.11}$$

Now $E = \bigcup_j E_j = \bigcup_{j,k} E_{j,k}$, and all the sets $E_{j,k}$ (over all j and k) are pairwise disjoint and bounded (each being contained in S_k), so the bounded case applies again to this doubly-indexed countable disjoint bounded family:

$$m(E) = \sum_{j,k} m(E_{j,k}) = \sum_{j=1}^{\infty} \left(\sum_{k=1}^{\infty} m(E_{j,k})\right) = \sum_{j=1}^{\infty} m(E_j). \tag{6.12}$$

■

Remark 6.1. The key point used in this argument was that each E_j could be approximated from inside by a *compact* set F_j. Disjointness of the E_j then translates into a positive distance between the compact pieces F_j, which is what makes the finite additivity of outer measure on compact sets available.

Corollary 6.3

If $F \subseteq E$ are both measurable with $m(F) < \infty$, then

$$m(E - F) = m(E) - m(F). \tag{6.13}$$

Proof. $E = F \cup (E - F)$, and they are disjoint. So by additivity,

$$m(E) = m(F) + m(E - F). \tag{6.14}$$

Since $m(F) < \infty$, we can subtract it from both sides and the desired equality follows. ∎

Proposition 6.4

Let $E \subseteq \mathbb{R}^d$ be Lebesgue measurable. Then the following hold.

(i) If $m(E) < \infty$, then for every $\varepsilon > 0$ there exists a compact set $K \subseteq E$ with

$$m(E - K) \leq \varepsilon. \tag{6.15}$$

(ii) If $m(E) < \infty$, then for every $\varepsilon > 0$ there exists a finite union $F = \bigcup_{i=1}^{N} Q_i$ of almost disjoint closed cubes such that

$$m(E \triangle F) \leq \varepsilon, \tag{6.16}$$

where $E \triangle F = (E - F) \cup (F - E)$ denotes the symmetric difference.

Proof. (i) Write E as the disjoint union of its intersections with the shells S_k from earlier:

$$E = \bigsqcup_{k=1}^{\infty} E_k, \qquad E_k = E \cap S_k. \tag{6.17}$$

By countable additivity, $m(E) = \sum_{k=1}^{\infty} m(E_k)$. Since $m(E) < \infty$ this series converges, so there is N such that

$$\sum_{k=1}^{N} m(E_k) \geq m(E) - \frac{\varepsilon}{2}. \tag{6.18}$$

Since $\bigcup_{k=1}^{N} S_k = B_n$ (the closed ball of radius N),

$$m(E \cap B_n) = \sum_{k=1}^{n} m(E_k) \;\geq\; m(E) - \frac{\varepsilon}{2}. \tag{6.19}$$

Now, by applying Lemma 6.1 to $E \cap B_n$, there is a closed set $K \subseteq E \cap B_n$ with

$$m((E \cap B_n) - K) \leq \frac{\varepsilon}{2}. \tag{6.20}$$

Since $K \subseteq B_n$, K is closed and bounded, hence compact, and $K \subseteq E$. Finally

$$E - K = (E - (E \cap B_n)) \cup ((E \cap B_n) - K), \tag{6.21}$$

a disjoint union, so

$$m(E - K) = m(E) - m(E \cap B_n) + m((E \cap B_n) - K) \;\leq\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{6.22}$$

(ii) Fix an open set $O \supseteq E$ with $m(O - E) < \varepsilon/2$. In particular, $m(O) < \infty$. Every open subset of $\mathbb{R}^d$ can be written as a countable union of almost disjoint closed cubes,

$$O = \bigcup_{j=1}^{\infty} Q_j, \qquad \{Q_j\} \text{ almost disjoint}, \tag{6.23}$$

and by countable additivity of almost-disjoint cubes (Proposition 5.10),

$$m(O) = \sum_{j=1}^{\infty} |Q_j| < \infty. \tag{6.24}$$

Choose N so that

$$\sum_{j=1}^{N} |Q_j| \;\geq\; m(O) - \frac{\varepsilon}{2}, \tag{6.25}$$

and set $F = \bigcup_{j=1}^{N} Q_j$, a finite union of almost disjoint closed cubes with $F \subseteq O$ and

$$m(O - F) = m(O) - m(F) \leq \frac{\varepsilon}{2}. \tag{6.26}$$

Now estimate the symmetric difference $E \triangle F = (E - F) \cup (F - E)$: Since $E \subseteq O$, we have $E - F \subseteq O - F$, so

$$m(E - F) \leq m(O - F) \leq \frac{\varepsilon}{2}. \tag{6.27}$$

Since $F \subseteq O$, we have $F - E \subseteq O - E$, so

$$m(F - E) \leq m(O - E) \leq \frac{\varepsilon}{2}. \tag{6.28}$$

Adding, $m(E \triangle F) \leq m(E - F) + m(F - E) \leq \varepsilon$.

■

Corollary 6.5

If $E \subseteq \mathbb{R}^d$ is measurable, then

$$m\,(E) = \sup\{m\,(K) : K \subseteq E \text{ and } K \text{ is compact }\}\,. \tag{6.29}$$

Proof. The case $m(E) < \infty$ follows trivially from Proposition 6.4(i). Now suppose $m(E) = \infty$.

For each n, consider $E_n = E \cap \overline{B(0,n)}$. Then $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ is an increasing sequence, and the union $\bigcup_n E_n = E$, so we have (by Proposition 1.4)

$$\lim_{n\to\infty} m(E_n) = m(E) = \infty. \tag{6.30}$$

For each n, $m(E_n) < \infty$, so we find compact sets $K_n \subseteq E_n$ such that $m(E_n - K_n) \leq 1$, i.e. $m(E_n) \leq m(K_n) + 1$. Since $m(E_n) \to \infty$, we have

$$\sup_n m(K_n) \geq \sup_n m(E_n) - 1 = \infty = m(E). \tag{6.31}$$

So we are done! ■

Remark 6.2. These statements express the same intuitive fact in different guises. A measurable set can be approximated from outside by an open set, from inside by a closed set, from inside by a compact set (when it has finite measure), and by a finite union of almost-disjoint cubes, up to a set of arbitrarily small measure. One chooses whichever form of approximation is most convenient for the problem at hand.

§6.ii. Completeness of Lebesgue measure

One can show that $(\mathbb{R}^d, \mathcal{L}, m)$ is a complete measure space.

Theorem 6.6

m is a complete measure.

Proof. Suppose $E \in \mathcal{L}$ with $m(E) = 0$. Then for any subset $A \subseteq E$, by the monotonicity of exterior measure, $m^*(A) \leq m^*(E) = 0$, implying that A is also a null-set. Since null-sets are measurable, $A \in \mathcal{L}$. So m is a complete measure. ■

The Lebesgue measurable sets arise as the completion of the σ-algebra of Borel sets, that is by adjoining all subsets of Borel sets of measure zero.

Definition 6.1. A set is called G_δ set if it is the intersection of a countable collection of open sets. A set is called F_σ set if it is the union of a countable collection of closed sets.

Proposition 6.7

Let $E \subseteq \mathbb{R}^d$. The following are equivalent:

(a) E is measurable.

(b) There is a G_δ-set $G \supseteq E$ such that $m^*(G - E) = 0$.

(c) There is a F_σ-set $F \subseteq E$ such that $m^*(E - F) = 0$.

Proof. **(b⇒a)** G is measurable, since its the intersection of countably many measurable sets. So is $G - E$, since it's a null-set. Therefore, $E = G - (G - E)$ is a measurable set.

(c⇒a) F is measurable, since its the union of countably many measurable sets. So is $E - F$, since it's a null-set. Therefore, $E = F \cup (E - F)$ is a measurable set.

(a⇒b) Since E is measurable, for every n, there exists an open set $O_n \supseteq E$ such that

$$m^*(O_n - E) \leq \frac{1}{n}. \tag{6.32}$$

Choose $G = \bigcap_{n=1}^{\infty} O_n \supseteq E$. Then,

$$G - E = \bigcap_{n=1}^{\infty} O_n - E \subseteq O_n - E \tag{6.33}$$

$$\implies m^*(G - E) \leq m^*(O_n - E) \leq \frac{1}{n}, \tag{6.34}$$

for every n. Therefore, $m^*(G - E) = 0$.

(a⇒c) Since E is measurable, for every n, there exists an open set $F_n \subseteq E$ such that

$$m^*(E - F_n) \leq \frac{1}{n}. \tag{6.35}$$

Choose $F = \bigcup_{n=1}^{\infty} F_n \subseteq E$. Then,

$$E - F = E - \bigcap_{n=1}^{\infty} F_n \subseteq E - F_n \tag{6.36}$$

$$\implies m^*(E - F) \leq m^*(E - F_n) \leq \frac{1}{n}, \tag{6.37}$$

for every n. Therefore, $m^*(E - F) = 0$. ■

Proposition 6.8

$E \subseteq \mathbb{R}^d$ is measurable if and only if for every $\varepsilon > 0$, there exists $F \subseteq E \subseteq U$ with U open and F closed such that $m^*(U - F) \leq \varepsilon$.

Proof. Suppose E is measurable. Then there exists $F \subseteq E \subseteq U$ with U open and F closed such that

$$m^*(U - E) \leq \frac{\varepsilon}{2} \text{ and } m^*(E - F) \leq \frac{\varepsilon}{2}. \tag{6.38}$$

Then $U - F = (U - E) \cup (E - F)$ is a disjoint union, and hence

$$m^*(U - F) \leq m^*(U - E) + m^*(E - F) \leq \varepsilon. \tag{6.39}$$

Conversely, suppose such $F \subseteq E \subseteq U$ exists for every ε. In particular, for every n, we can choose

$F_n \subseteq E \subseteq U_n$ with F_n closed and U_n open, and

$$m^*(U_n - F_n) \leq \frac{1}{n}. \tag{6.40}$$

Choose $F = \bigcup_{n=1}^{\infty} F_n$ is a F_σ set. Then we have $F \subseteq E$. Furthermore,

$$E - F = E - \bigcup_{n=1}^{\infty} F_n \subseteq E - F_n \subseteq U_n - F_n. \tag{6.41}$$

By monotonicity, we have

$$m^*(E - F) \leq m^*(U_n - F_n) \leq \frac{1}{n}, \tag{6.42}$$

which is true for every n. Therefore, $m^*(E - F) = 0$. This gives E is measurable by Proposition 6.7. ■

§6.iii. Invariance properties of Lebesgue measure

Now we'll discuss some invariance properties of Lebesgue measure.

Theorem 6.9 (Translation invariance)

Let $E \subseteq \mathbb{R}^d$ be measurable and $x_0 \in \mathbb{R}^d$. Then $E + x_0 := \{x + x_0 : x \in E\}$ is measurable and

$$m(E + x_0) = m(E). \tag{6.43}$$

Proof. We divide the proof into several steps.

Claim 6.1: Outer measure is translation invariant.

Proof. Every covering of E by cubes $\{Q_j\}$ gives, upon translating each cube by x_0, a covering of $E + x_0$ by cubes $\{Q_j + x_0\}$ of the *same* volumes (translating a cube does not change its side lengths). Hence

$$m^*(E + x_0) = m^*(E) \tag{6.44}$$

for every set E, since the coverings of E and of $E + x_0$ by cubes are in volume-preserving bijection with each other. □

Claim 6.2: If E is measurable, so is $E + x_0$.

Proof. If E is measurable, for $\varepsilon > 0$ choose an open $O \supseteq E$ with $m^*(O - E) \leq \varepsilon$. Then $O + x_0$ is open, contains $E + x_0$, and

$$(O + x_0) - (E + x_0) = (O - E) + x_0, \tag{6.45}$$

so by Claim 6.1,

$$m^*((O + x_0) - (E + x_0)) = m^*(O - E) \leq \varepsilon. \tag{6.46}$$

□

Hence $E + x_0$ is measurable, and its measure equals $m^*(E + x_0) = m^*(E) = m(E)$. ■

Theorem 6.10 (Scaling/dilation invariance)

Let $E \subseteq \mathbb{R}^d$ be measurable and $\lambda > 0$. Then $\lambda E := \{\lambda x : x \in E\}$ is measurable and

$$m(\lambda E) = \lambda^d \, m(E). \tag{6.47}$$

Proof. As in the translation case, we divide the proof into several steps.

Claim 6.3: $m^*(\lambda E) = \lambda^d \, m^*(E)$ for every $E \subseteq \mathbb{R}^d$.

Proof. If $\{Q_j\}$ is a covering of E by cubes, then $\{\lambda Q_j\}$ covers λE, and a cube of side length ℓ scales to a cube of side length $\lambda \ell$, so its volume scales by λ^d. Hence

$$m^*(\lambda E) = \lambda^d \, m^*(E), \tag{6.48}$$

for every set E, since the coverings of E and of λE by cubes are in volume-scaling-by-λ^d bijection with each other. □

Claim 6.4: If E is measurable, so is λE.

Proof. If E is measurable, choosing an open $O \supseteq E$ with $m^*(O - E) < \frac{\varepsilon}{\lambda^d}$. The set λO is open, contains λE, and $(\lambda O) - (\lambda E) = \lambda(O - E)$, so

$$m^*(\lambda O - \lambda E) = \lambda^d \, m^*(O - E) \leq \varepsilon, \tag{6.49}$$

as desired. □

Hence, λE is measurable with $m(\lambda E) = \lambda^d m(E)$. ■

One can similarly prove a rotation invariance. However one can prove a more general fact.

§6.iv. Lebesgue measure and linear maps

Theorem 6.11

Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a linear map. If $E \subseteq \mathbb{R}^d$ is measurable, so is $T(E)$. Furthermore,

$$m(T(E)) = |\det T| \, m(E). \tag{6.50}$$

We shall prove it in several steps.

Proposition 6.12

Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be a linear map with $\det T = 0$. Then $m^*\left(T\left(\mathbb{R}^d\right)\right) = 0$.

Proof. Let $H = T\left(\mathbb{R}^d\right)$. We can write

$$H = \bigcup_{R=1}^{\infty} \left(H \cap [-R, R]^d\right) =: \bigcup_{R=1}^{\infty} H_R. \tag{6.51}$$

By countable additivity,

$$m^*(H) \leq \sum_{R=1}^{\infty} m^*\left(H \cap [-R, R]^d\right). \tag{6.52}$$

So we will prove that $m^*\left(H \cap [-R, R]^d\right) = 0$ for every R. Fix N, and let's divide each side of $[-R, R]^d$ into N equal parts, i.e. divide the cube into N^d small cubes, each of side length $\frac{2R}{N}$. Let

$$\mathcal{Q} = \left\{Q : Q \text{ is a small cube that intersects } H \cap [-R, R]^d\right\}. \tag{6.53}$$

Then $\mathcal{Q}$ covers $H \cap [-R, R]^d$, so

$$m^*\left(H \cap [-R, R]^d\right) \leq \sum_{Q \in \mathcal{Q}} |Q| = |\mathcal{Q}| \left(\frac{2R}{N}\right)^d, \tag{6.54}$$

where $|\mathcal{Q}|$ denotes the size of the set $\mathcal{Q}$. [1]

Claim 6.5: $|\mathcal{Q}| \leq C \cdot N^{d-1}$ for some constant C (not depending on N).

Proof. Since H is a stricly lower-dimensional hyperplane through the origin, there exists some nonzero vector $\mathbf{v} \in \mathbb{R}^d$ such that $\mathbf{x} \cdot \mathbf{v} = 0$ for every $\mathbf{x} \in H$. WLOG we can assume $v_d \neq 0$ (well, there is a coordinate $v_{i_0} \neq 0$, then d can be replaced with i_0 the following argument). Consider a "column"

$$C = I_1 \times I_2 \times \cdots \times I_{d-1} \times [-R, R], \tag{6.55}$$

where I_i are intervals of length $\frac{2R}{N}$ in each direction. There are N^{d-1} such columns. Therefore,

$$|\mathcal{Q}| \leq N^{d-1} \times (\text{maximum number of cubes in a column that intersect } H_R). \tag{6.56}$$

[1] **Abuse of notation:** For a cube Q, $|Q|$ is its volume, and for the set $\mathcal{Q}$, $|\mathcal{Q}|$ is its cardinality.

If $\mathbf{y}, \mathbf{z} \in H_R$ that belong to the same column, then

$$\mathbf{y} \cdot \mathbf{v} = 0 \implies y_d = -\sum_{i=1}^{d-1} \frac{v_i}{v_d} y_i, \tag{6.57}$$

$$\mathbf{z} \cdot \mathbf{v} = 0 \implies z_d = -\sum_{i=1}^{d-1} \frac{v_i}{v_d} z_i, \tag{6.58}$$

$$\implies |y_d - z_d| \leq \left| \sum_{i=1}^{d-1} \frac{v_i}{v_d} (y_i - z_i) \right| \leq \sum_{i=1}^{d-1} \left| \frac{v_i}{v_d} \right| \frac{2R}{N}, \tag{6.59}$$

since $\mathbf{y}$ and $\mathbf{z}$'s i-th coordinate differs by at most $\frac{2R}{N}$. Hence, if $\mathbf{y}, \mathbf{z} \in H_R$ that belong to the same column, then their d-th coordinate differs by at most $M \cdot \frac{2R}{N}$, for some constant M that doesn't depend on N (but M may depend on T).

Therefore, we can argue that there are **at most** $M+2$ cubes in a column that intersect H_R. Otherwise, we can find two cubes in a column, both of them intersect H_R, and there are at least $M+1$ cubes in between them. This way, the distance between the d-th coordinate of those two points (at which the said cubes intersect H_R) would be at least $(M+1)\frac{2R}{N}$.

Hence, $|\mathcal{Q}| \leq (M+2)\, N^{d-1}$. □

Now, using (6.54),

$$m^* \left(H \cap [-R, R]^d \right) \leq |\mathcal{Q}| \left(\frac{2R}{N} \right)^d = C \cdot N^{d-1} \frac{(2R)^d}{N^d} = \frac{C'}{N}. \tag{6.60}$$

This is true for every N. By taking $N \to \infty$, we get $m^*(H_R) = 0$. By using subadditivity (6.52), we get that $m^*(H) = 0$. ■

This proves the $\det T = 0$ case of Theorem 6.11. Now, we focus on the $\det T \neq 0$ case. A standard linear algebra fact says that any non-singular linear map can be written as a composition of "elementary" linear maps of the following types:

(a) $T_1(x_1, \ldots, x_i, \ldots, x_d) = (x_1, \ldots, cx_i, \ldots, x_d)$ for $c \neq 0$;

(b) $T_2(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_d) = (x_1, \ldots, x_j, \ldots, x_i, \ldots, x_d)$.

(c) $T_3(x_1, x_2 \ldots, x_d) = (x_1 + x_2, x_2, \ldots, x_d)$.

Also, note that

$$\det T_1 = c, \quad \det T_2 = -1, \quad \det T_3 = 1. \tag{6.61}$$

Lemma 6.13

Let $Q = [0,1]^d$ be the unit cube. Then

(a) $m^*(T_1(Q)) = |c|$;

(b) $m^*(T_2(Q)) = 1$;

(c) $m^*(T_3(Q)) \leq 1$.

In other words, $m^*(T_i(Q)) \leq |\det T_i|$ for each $i = 1, 2, 3$.

Proof. For T_1, the image $T_1(Q)$ is the rectangle

$$[0,1] \times \cdots \times [0,1] \times [0,c] \times [0,1] \times \cdots \times [0,1], \tag{6.62}$$

or $[c,0]$ instead of $[0,c]$. We know that the exterior measure of a rectangle is its volume (Proposition 5.6), hence,

$$m^*(T_1(Q)) = |c| = |\det T_1|. \tag{6.63}$$

For T_2, the image $T_2(Q) = Q$, so

$$m^*(T_2(Q)) = m^*(Q) = 1 = |\det T_2|. \tag{6.64}$$

For T_3, note that

$$T_3(Q) = \{(x_1, \ldots, x_d) : 0 \leq x_i \leq 1 \text{ for } i \geq 2; \text{ and } x_2 \leq x_1 \leq x_2 + 1\}. \tag{6.65}$$

Note that $\widetilde{Q} := T_3(Q)$ is contained in the rectangle $[0,2] \times [0,1]^{d-1}$. We cover this rectangle using cubes of side length $\frac{1}{N}$. Let

$$\mathcal{Q} = \left\{Q' : Q' \text{ is a small cube that intersects } \widetilde{Q}\right\}. \tag{6.66}$$

Then $\mathcal{Q}$ covers $\widetilde{Q}$, so

$$m^*\left(\widetilde{Q}\right) \leq \sum_{Q' \in \mathcal{Q}} |Q'| = |\mathcal{Q}| \cdot \frac{1}{N^d}. \tag{6.67}$$

Again, $|\mathcal{Q}|$ denotes the size of the set $\mathcal{Q}$.

Claim 6.6: $|\mathcal{Q}| = (1+N)\, N^{d-1}$.

Proof. Fix a "row"

$$R = [0,2] \times I_2 \times I_3 \times \cdots \times I_d, \tag{6.68}$$

where each I_i is an interval of length $\frac{1}{N}$. There are N^{d-1} such rows. Now let's see how many small cubes of this row intersect $\widetilde{Q}$.

Suppose $I_2 = \left[\frac{k}{N}, \frac{k+1}{N}\right]$ for some k. Now, take $\mathbf{x} \in \widetilde{Q} \cap R$. Then

$$x_2 \leq x_1 \leq x_2 + 1 \implies x_1 \in \left[\frac{k}{N}, \frac{k+(N+1)}{N}\right]. \tag{6.69}$$

Therefore, a small cube $Q' = I_1 \times I_2 \times \cdots \times I_d$ intersects $\widetilde{Q}$ if and only if $I_1 \subseteq \left[\frac{k}{N}, \frac{k+(N+1)}{N}\right]$. There are $N+1$ choices for I_1 then. As a result, there are $(N+1)$ cubes in R that intersect

$\widetilde{Q}$. Since this is the same for all rows, we have $|\mathcal{Q}| = (1+N)N^{d-1}$. □

Now, (6.67) gives us

$$m^*\left(T_3\left(Q\right)\right) \leq \frac{(1+N)N^{d-1}}{N^d} = 1 + \frac{1}{N}. \tag{6.70}$$

Since this is true for every N, by taking $N \to \infty$, we get $m^*\left(T_3\left(Q\right)\right) \leq 1$. ■

So we have shown that under elementary linear maps T, we have

$$m^*\left(T\left(Q\right)\right) \leq \left|\det T\right| m^*\left(Q\right), \tag{6.71}$$

where $Q = [0,1]^d$ is the unit cube.

Lemma 6.14

For **every** cube Q' and elementary linear map T,

$$m^*\left(T\left(Q'\right)\right) \leq \left|\det T\right| m^*\left(Q'\right). \tag{6.72}$$

Proof. If Q' is another cube, then it's just a dilation and translation of Q, i.e. $Q' = \delta Q + s$. Then

$$T\left(Q'\right) = T\left(\delta Q + s\right) = \delta\, T\left(Q\right) + T\left(s\right). \tag{6.73}$$

By the dilation and translation invariance of outer measure,

$$\begin{aligned} m^*\left(T\left(Q'\right)\right) &= m^*\left(\delta\, T\left(Q\right) + T\left(s\right)\right) = \delta^d\, m^*\left(T\left(Q\right)\right) && (6.74)\\ &\leq \left|\det T\right| \delta^d\, m^*\left(Q\right) = \left|\det T\right| m^*\left(\delta Q + s\right) && (6.75)\\ &= \left|\det T\right| m^*\left(Q'\right). && (6.76) \end{aligned}$$

■

Lemma 6.15

For **every** open set $O \subseteq \mathbb{R}^d$ and elementary linear map T,

$$m^*\left(T\left(O\right)\right) \leq \left|\det T\right| m^*\left(O\right). \tag{6.77}$$

Proof. Write O as a almost-disjoint coutable union of cubes: $O = \bigcup_{i=1}^{\infty} Q_i$. Then

$$\begin{aligned} m^*\left(T\left(O\right)\right) &= m^*\left(T\left(\bigcup_{i=1}^{\infty} Q_i\right)\right) = m^*\left(\bigcup_{i=1}^{\infty} T\left(Q_i\right)\right) && (6.78)\\ &\leq \sum_{i=1}^{\infty} m^*\left(T\left(Q_i\right)\right) && (6.79)\\ &\leq \left|\det T\right| \sum_{i=1}^{\infty} m^*\left(Q_i\right) = \left|\det T\right| m^*\left(O\right), && (6.80) \end{aligned}$$

where the last equality follows from Proposition 5.10. ■

Lemma 6.16

Let $G \subseteq \mathrm{GL}(d, \mathbb{R})$ be the set

$$G = \left\{ T \in \mathrm{GL}(d, \mathbb{R}) \ : \ m^*(T(O)) \leq |\det T| \ m^*(O) \ \text{ for every open } O \subseteq \mathbb{R}^d \right\}. \tag{6.81}$$

Then G is closed under composition.

Proof. A linear map is continuous, and hence, an invertible linear map is a homeomorphism $\mathbb{R}^d \to \mathbb{R}^d$. Therefore, if O is open and $S, T \in G$, then both $T(O)$ and $ST(O)$ are open.

Now, since $T(O)$ is open and $S \in G$, we have

$$m^*(ST(O)) \leq |\det S| \, m^* T(O). \tag{6.82}$$

Since $T \in G$ and O is open, we have

$$m^*(T(O)) \leq |\det T| \ m^*(O). \tag{6.83}$$

Combining these, we have

$$m^*(ST(O)) \leq |\det S| \, |\det T| \ m^*(O) = |\det(ST)| \, m^*(O), \tag{6.84}$$

implying that $ST \in G$. ■

Proposition 6.17

$G = \mathrm{GL}(d, \mathbb{R})$.

Proof. Every invertible linear map $T : \mathbb{R}^d \to \mathbb{R}^d$ can be written as a composition of the elementary maps. By Lemma 6.15, elementary maps are in G. Since G is closed under composition, we have $T \in G$ for every T. ■

In other words, for **every** $T \in \mathrm{GL}(d, \mathbb{R})$, we have

$$m^*(T(O)) \leq |\det T| \, m^*(O). \tag{6.85}$$

Actually, we can prove the equality in (6.85).

Proposition 6.18

For every open set $O \subseteq \mathbb{R}^d$ and $T \in \mathrm{GL}(d, \mathbb{R})$, we have

$$m^*(T(O)) = |\det T| \, m^*(O). \tag{6.86}$$

Proof. We already have $m^*(T(O)) \leq |\det T|\, m^*(O)$. Now note that

$$m^*(O) = m^*\left(T^{-1}(T(O))\right). \tag{6.87}$$

Since $T(O)$ is also open, and $T^{-1} \in \mathrm{GL}(d, \mathbb{R})$, using (6.85),

$$m^*(O) = m^*\left(T^{-1}(T(O))\right) \leq \left|\det\left(T^{-1}\right)\right| m^*(T(O)) = \frac{1}{|\det T|} m^*(T(O)). \tag{6.88}$$

Rearranging, we get

$$m^*(T(O)) \geq |\det T|\, m^*(O). \tag{6.89}$$

■

Proposition 6.19

For every $E \subseteq \mathbb{R}^d$,

$$m^*(T(E)) = |\det T|\, m^*(E). \tag{6.90}$$

Proof. We know from Proposition 5.7 that

$$m^*(E) = \inf_{O \supseteq E, O \text{ open}} m^*(O). \tag{6.91}$$

Since T is a homeomorphism, $O \supseteq T(E)$ if and only if $O = T(U)$ for some open $U \supseteq E$.

$$m^*(T(E)) = \inf_{\substack{O \supseteq T(E), \\ O \text{ open}}} m^*(O) = \inf_{\substack{U \supseteq E, \\ U \text{ open}}} m^*(T(U)) \tag{6.92}$$

$$= |\det T| \inf_{\substack{U \supseteq E, \\ U \text{ open}}} m^*(U) = |\det T|\, m^*(E). \tag{6.93}$$

■

Now we are pretty much ready to finish the proof of Theorem 6.11.

Proof of Theorem 6.11. Suppose E is measurable, and $\varepsilon > 0$. Then we can find an open set $O \supseteq E$ such that

$$m^*(O - E) \leq \frac{\varepsilon}{\det T}. \tag{6.94}$$

Since T is a homeomorphism, $T(O)$ is an open set containing $T(E)$, and $T(O) - T(E) = T(O - E)$. As a result,

$$m^*(T(O) - T(E)) = m^*(T(O - E)) = |\det T|\, m^*(O - E) \leq \varepsilon. \tag{6.95}$$

So $T(E)$ is measurable. Then $m(T(E)) = |\det T|\, m(E)$ follows from Proposition 6.19. ■

Remark 6.3. For a linear map $T : \mathbb{R}^d \to \mathbb{R}^d$, measurability of E implies the measurability of $T(E)$ because we are working in the same dimension. If $T : \mathbb{R}^d \to \mathbb{R}^{d'}$ for $d' < d$, then

measurability of E does not always imply the measurability of $T(E)$. For instance, suppose $T : \mathbb{R}^2 \to \mathbb{R}^1$ be the projection onto the first component.

Consider $E = A \times \{0\}$, where A is the non-measurable set constructed in Theorem 3.11. Then E has outer measure 0 as it can be covered using the 0-measure rectangle $[0,1] \times [0,0]$. So $E \subseteq \mathbb{R}^2$ is measurable. But $T(E) = A \subset \mathbb{R}$ is **NOT** measurable.

§6.v. Exercises

Exercise 6.1. Find a sequence of real-valued non-negative functions $\{f_k\}_{k\geq 1}$ on $[0,1]$ so that

$$\limsup_{k\to\infty} f_k(x) = \infty \ \forall x \in [0,1] \quad \text{and} \quad \int_0^1 f_k(x)\,\mathrm{d}x \to 0. \tag{6.96}$$

Solution. Define

$$\begin{gathered} f_1 = \chi_{[0,1]} \\ f_2 = 2\chi_{[0,\frac{1}{2}]},\ f_3 = 2\chi_{[\frac{1}{2},1]} \\ f_4 = 3\chi_{[0,\frac{1}{4}]},\ f_5 = 3\chi_{[\frac{1}{4},\frac{1}{2}]},\ f_6 = 3\chi_{[\frac{1}{2},\frac{3}{4}]},\ f_7 = 3\chi_{[\frac{3}{4},1]} \\ \cdots \end{gathered}$$

In other words, for $n = 2^k + m$, with k being the unique number satisfying $2^k \leq n < 2^{k+1}$, and $0 \leq m < 2^k$, we define

$$f_n = (k+1)\,\chi_{[\frac{m}{2^k},\frac{m+1}{2^k}]}. \tag{6.97}$$

Then

$$\int_0^1 f_n(x)\,\mathrm{d}x = \frac{k+1}{2^k} \to 0. \tag{6.98}$$

Now, given any $x \in [0,1]$, and for every $M > 0$, there are non-negative integers k, m such that

$$k > M \text{ and } m \leq 2^k x \leq m+1. \tag{6.99}$$

Now, $f_{2^k+m}(x) = k+1 > M$. Therefore, $\limsup_{k\to\infty} f_k(x) = \infty$. ■

Exercise 6.2. Let $f : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Assume that there exists a positive constant c with the following property: If $A \subseteq \mathbb{R}$ is bounded and measurable, and f is bounded on A, then

$$\left| \int_A f(x)\,\mathrm{d}x \right| \leq c. \tag{6.100}$$

Prove or disprove: f is Lebesgue integrable.

Solution. Yes, f is Lebesgue integrable. We shall prove it by showing that $\int_{\mathbb{R}} f^+\,\mathrm{d}m \leq c$ and $\int_{\mathbb{R}} f^-\,\mathrm{d}m \leq c$. Given $n, M \geq 1$, we define

$$A_{n,M} = \{x \in [-n,n] : 0 \leq f(x) \leq M\}. \tag{6.101}$$

Then A is bounded, measurable (since it's the intersection of $[-n,n]$ and $\{0 \le f \le M\}$), and f is bounded on A. Therefore,

$$\left|\int_{A_{n,M}} f(x)\,\mathrm{d}x\right| \le c \implies \int_{A_{n,M}} f^+(x)\,\mathrm{d}x \le c. \tag{6.102}$$

By Corollary 2.12, since $\{A_{n,M}\}$ is an increasing sequence of measurable sets, and

$$A_M = \bigcup_{n=1}^{\infty} A_{n,M} = \{x \in \mathbb{R} : 0 \le f(x) \le M\}. \tag{6.103}$$

Therefore,

$$\int_{A_M} f^+(x)\ \mathrm{d}x = \lim_{n\to\infty} \int_{A_{n,M}} f^+(x)\,\mathrm{d}x \le c. \tag{6.104}$$

Now we again apply Corollary 2.12 on the increasing sequence $\{A_M\}_M$.

$$\bigcup_{M=1}^{\infty} A_M = \{x \in \mathbb{R} : f(x) \ge 0\}. \tag{6.105}$$

Now,

$$\int_{\mathbb{R}} f^+(x)\,\mathrm{d}x = \int_{\bigcup_{M=1}^{\infty} A_M} f^+(x)\,\mathrm{d}x = \lim_{M\to\infty} \int_{A_M} f^+(x)\ \mathrm{d}x \le c. \tag{6.106}$$

Similarly, doing the same procedure on the function $-f$, we will get $\int_{\mathbb{R}} f^-(x)\,\mathrm{d}x \le c$. Therefore, f is Lebesgue-integrable. ■

Exercise 6.3. Let $A \subset [0,1]$ be a closed set. For each positive integer N let $b(N)$ be the number of integers n such that $0 \le n \le N-1$ and

$$\left[\frac{n}{N}, \frac{n+1}{N}\right] \cap A \ne \varnothing. \tag{6.107}$$

Show that

$$\lim_{N\to\infty} \frac{b(N)}{N} = m(A), \tag{6.108}$$

where m denotes the Lebesgue measure.

Solution. Let B_N be the union of the intervals $\left[\frac{n}{N}, \frac{n+1}{N}\right]$ that intersect A.

$$B_N = \bigcup_{\left[\frac{n}{N}, \frac{n+1}{N}\right] \cap A \ne \varnothing} \left[\frac{n}{N}, \frac{n+1}{N}\right] \tag{6.109}$$

Since $A \subset [0,1]$, $A \subseteq B_N$. Therefore,

$$m(A) \le m(B_N) \le \sum_{\left[\frac{n}{N}, \frac{n+1}{N}\right] \cap A \ne \varnothing} m\left(\left[\frac{n}{N}, \frac{n+1}{N}\right]\right) = \frac{b(N)}{N}. \tag{6.110}$$

Therefore,

$$m(A) \leq \liminf_{n\to\infty} \frac{b(N)}{N}. \tag{6.111}$$

Now, let $c(N) = N - b(N)$ be the number of intervals $\left[\frac{n}{N}, \frac{n+1}{N}\right]$ that don't intersect A, i.e. intervals that are contained in $A^c = [0,1] - A$. Given $\varepsilon > 0$, we can choose a compact set $K \subseteq A^c$ such that

$$m(A^c) - m(K) = m(A^c - K) \leq \varepsilon. \tag{6.112}$$

Since K is compact and A is closed, and they are disjoint, by Claim 5.2, $d(A,K) = \delta > 0$. Choose N such that $\delta N > 1$. Now, let

$$C_N = \bigcup_{\left[\frac{n}{N}, \frac{n+1}{N}\right] \subseteq A^c} \left[\frac{n}{N}, \frac{n+1}{N}\right]. \tag{6.113}$$

Claim 6.7: $K \subseteq C_N$.

Proof. Take $x \in K$. Then x is in some interval of the form $\left[\frac{n}{N}, \frac{n+1}{N}\right]$. If this interval intersects A at a, then

$$|a - x| \leq \frac{1}{N} < \delta = d(A,K), \tag{6.114}$$

which is a contradiction. Therefore, $\left[\frac{n}{N}, \frac{n+1}{N}\right] \subseteq A^c$. So this interval is in C_N. Hence, every $x \in K$ is contained in C_N. □

Therefore,

$$m(K) \leq m(C_N) \leq \sum_{\left[\frac{n}{N}, \frac{n+1}{N}\right] \subseteq A^c} m\left(\left[\frac{n}{N}, \frac{n+1}{N}\right]\right) = \frac{c(N)}{N}. \tag{6.115}$$

Combining this with $m(K) \geq m(A^c) - \varepsilon$, we get

$$m(A^c) - \varepsilon \leq \frac{c(N)}{N} \implies 1 - m(A) - \varepsilon \leq 1 - \frac{b(N)}{N} \implies \frac{b(N)}{N} \leq m(A) + \varepsilon, \tag{6.116}$$

for every ε and for some sufficiently large N. Therefore, taking $\varepsilon \to 0^+$, we have

$$\limsup_{n\to\infty} \frac{b(N)}{N} \leq m(A). \tag{6.117}$$

Combining (6.111) and (6.117), we can conclude that $\lim_{n\to\infty} \frac{b(N)}{N}$ exists, and it equals $m(A)$. ■

Note that for Exercise 6.3 to hold, it's necessary to have A closed. If A is not closed, $\frac{b(N)}{N} \to m(A)$ doesn't hold. For example, we can take

$$A = \mathbb{Q} \cap [0,1], \tag{6.118}$$

which is not closed. For this A, $b(N) = N$ always, so we have

$$\lim_{N\to\infty} \frac{b(N)}{N} = 1 \neq 0 = m(A). \tag{6.119}$$

However, it **IS** true that

$$m\left(\overline{A}\right) = \lim_{N\to\infty} \frac{b(N)}{N}. \tag{6.120}$$

First, let $b_{\overline{A}}(N)$ be the number of integers $0 \leq n \leq N-1$ such that

$$\left[\frac{n}{N}, \frac{n+1}{N}\right] \cap \overline{A} \neq \varnothing. \tag{6.121}$$

Then $b_{\overline{A}}(N) \geq b(N)$, since an interval intersects A implies it intersects $\overline{A}$ as well. Therefore,

$$m\left(\overline{A}\right) = \lim_{N\to\infty} \frac{b_{\overline{A}}(N)}{N} = \limsup_{N\to\infty} \frac{b_{\overline{A}}(N)}{N} \geq \limsup_{N\to\infty} \frac{b(N)}{N}. \tag{6.122}$$

As before, define

$$B_N = \bigcup_{\left[\frac{n}{N}, \frac{n+1}{N}\right] \cap A \neq \varnothing} \left[\frac{n}{N}, \frac{n+1}{N}\right] \tag{6.123}$$

B_N is a closed set that contains A, therefore it contains $\overline{A}$ as well! Hence,

$$m\left(\overline{A}\right) \leq m\left(B_N\right) = \frac{b(N)}{N}. \tag{6.124}$$

Hence,

$$m\left(\overline{A}\right) \leq \liminf_{N\to\infty} \frac{b(N)}{N}. \tag{6.125}$$

Hence,

$$m\left(\overline{A}\right) = \lim_{N\to\infty} \frac{b(N)}{N}. \tag{6.126}$$

Exercise 6.4. For $n = 1, 2, \ldots$, let $f_n : \mathbb{R} \to \mathbb{R}$ be a Lebesgue measurable function. Assume that the set $A_n = \{x \in \mathbb{R} : |f_n(x)| > n^{-2}\}$ has Lebesgue measure $m(A_n) \leq 2^{-n}$. Show that the series $\sum_{n=1}^{\infty} f_n(x)$ is absolutely convergent for almost every $x \in \mathbb{R}$.

Solution. We have that

$$\sum_{n=1}^{\infty} m(A_n) \leq \sum_{n=1}^{\infty} 2^{-n} = 1 < \infty. \tag{6.127}$$

So we can apply Borel–Cantelli lemma. Let $E = \{x : x \in A_n \text{ for infinitely many } n\}$. By Borel–Cantelli lemma, $m(E) = 0$. We'll prove that $\sum_{n=1}^{\infty} |f_n(x)|$ converges for $x \notin E$.

For $x \notin E$, there exists some N such that $x \notin A_n$ for $n > N$. Therefore, for $n > N$, $|f_n(x)| \leq n^{-2}$.

$$\sum_{n=1}^{\infty} |f_n(x)| = \sum_{n=1}^{N} |f_n(x)| + \sum_{n=N+1}^{\infty} |f_n(x)| \leq \sum_{n=1}^{N} |f_n(x)| + \sum_{n=N+1}^{\infty} \frac{1}{n^2}. \tag{6.128}$$

The first is a finite sum, and the second sum converges. Therefore, $\sum_{n=1}^{\infty} |f_n(x)|$ converges. ■

Exercise 6.5. Let $F = (f_1, f_2, \ldots, f_d) : \mathbb{R}^d \to \mathbb{R}^d$ be a polynomial map, i.e., each $f_i \in \mathbb{R}[x_1, x_2, \ldots, x_d]$ for $i = 1, 2, \ldots, d$. Show that if $E \subset \mathbb{R}^d$ has Lebesgue measure zero, then $F(E)$ also has Lebesgue measure zero.

(Hint: You may first try to prove that a Lipschitz map $G : \mathbb{R}^d \to \mathbb{R}^d$ takes sets of measure zero into sets of measure zero.)

Solution. Suppose $G : U \subset \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz, i.e. there is a constant $C > 0$ such that

$$\|G(\mathbf{x}) - G(\mathbf{y})\| \leq C \cdot \|\mathbf{x} - \mathbf{y}\|, \tag{6.129}$$

for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. Let $E \subseteq \mathbb{R}^d$ be a measure-0 set. Then given any $\varepsilon > 0$, there exists a countable collection of cubes $\{Q_i\}_{i \geq 1}$ such that $E \subseteq \bigcup_{i=1}^{\infty} Q_i$ and $\sum_{i=1}^{\infty} |Q_i| \leq \frac{\varepsilon}{C^d d^{d/2}}$.

Now let's examine $G(Q_i)$. Suppose Q_i has side length ℓ_i. If we take any two points $G(\mathbf{x}), G(\mathbf{y}) \in G(Q_i)$,

$$\left|(G(\mathbf{x}) - G(\mathbf{y}))_j\right| \leq \|G(\mathbf{x}) - G(\mathbf{y})\| \leq C \cdot \|\mathbf{x} - \mathbf{y}\| \leq C\sqrt{d}\ell_i. \tag{6.130}$$

Therefore, $G(Q_i)$ is contained in a cube $\widetilde{Q_i}$ of side length $C\sqrt{d}\ell_i$. As a result,

$$G(E) \subseteq G\left(\bigcup_{i=1}^{\infty} Q_i\right) = \bigcup_{i=1}^{\infty} G(Q_i) \subseteq \bigcup_{i=1}^{\infty} \widetilde{Q_i}. \tag{6.131}$$

Now,

$$\sum_{i=1}^{\infty} \left|\widetilde{Q_i}\right| = \sum_{i=1}^{\infty} \left(C\sqrt{d}\ell_i\right)^d = C^d d^{d/2} \sum_{i=1}^{\infty} \ell_i^d = C^d d^{d/2} \sum_{i=1}^{\infty} |Q_i| \leq \varepsilon. \tag{6.132}$$

Therefore, $G(E)$ has measure 0, i.e. Lipschitz function takes measure-0 sets to measure-0 sets.

Now we move back to our problem. Let E be a set of measure 0. We write

$$E = \bigcup_{n=1}^{\infty} (E \cap B_n) =: \bigcup_{n=1}^{\infty} E_n, \quad \text{where } B_n = \overline{B(\mathbf{0}, n)}. \tag{6.133}$$

Since E has measure 0, by monotonicity, $m(E_n) = 0$. Now, F is a polynomial map, so $\|DF\|$ is continuous. Hence, it attains a maximum value on the compact set B_n, say K_n. As a result, on B_n, by Mean-Value inequality,

$$\|F(\mathbf{x}) - F(\mathbf{y})\| \leq K_n \|\mathbf{x} - \mathbf{y}\|. \tag{6.134}$$

Therefore, F is Lipschitz on B_n, and hence $F(E_n)$ has measure 0. Now,

$$m(F(E)) = m\left(\bigcup_{n=1}^{\infty} F(E_n)\right) \leq \sum_{n=1}^{\infty} m(F(E_n)) = 0. \tag{6.135}$$

■

Exercise 6.6. Let E_k be a sequence of Lebesgue measurable sets in $\mathbb{R}^d$ such that $\sum_{k=1}^{\infty} m(E_k) < \infty$. Let F_j be the subset of $\mathbb{R}^d$ consisting of the points that belong to precisely j of the subsets $\{E_k\}_{k=1}^{\infty}$. Show that

$$\sum_{k=1}^{\infty} m(E_k) = \sum_{j=1}^{\infty} j\, m(F_j). \tag{6.136}$$

Solution. We shall use the fact that $m(A) = \int_{\mathbb{R}^d} \chi_A \, \mathrm{d}x$. Then

$$\sum_{k=1}^{\infty} m(E_k) = \sum_{k=1}^{\infty} \int_{\mathbb{R}^d} \chi_{E_k} \, \mathrm{d}x = \int_{\mathbb{R}^d} \left(\sum_{k=1}^{\infty} \chi_{E_k} \right) \mathrm{d}x, \tag{6.137}$$

where the swapping is sum and integral is justified by Proposition 2.8. Since $\sum_{k=1}^{\infty} m(E_k) < \infty$, by Borel–Cantelli lemma, almost every element of $\mathbb{R}^d$ belong to only finitely many E_k's. In other words, $\sum_{k=1}^{\infty} \chi_{E_k}$ is finite almost everywhere.

Now, if $x \in F_j$, then x belongs to j-many elements in $\{E_k\}_{k\geq 1}$. Therefore,

$$\sum_{k=1}^{\infty} \chi_{E_k}(x) = j, \qquad \text{for } x \in F_j. \tag{6.138}$$

$\bigcup_{j=1}^{\infty} F_j$ is a disjoint union, which covers almost everywhere. Therefore,

$$\sum_{k=1}^{\infty} \chi_{E_k} = \sum_{j=1}^{\infty} j \chi_{F_j}, \qquad \text{a.e.} \tag{6.139}$$

Therefore, their integrals are the same, i.e.

$$\begin{aligned} \sum_{k=1}^{\infty} m(E_k) &= \int_{\mathbb{R}^d} \left(\sum_{k=1}^{\infty} \chi_{E_k} \right) \mathrm{d}x = \int_{\mathbb{R}^d} \left(\sum_{j=1}^{\infty} j \chi_{F_j} \right) \mathrm{d}x \\ &= \sum_{j=1}^{\infty} j \int_{\mathbb{R}^d} \chi_{F_j} \, \mathrm{d}x = \sum_{j=1}^{\infty} j m(F_j). \end{aligned} \tag{6.140}$$

Again, the swapping of sum and integral is fascilitated by Proposition 2.8. ■

Exercise 6.7. Construct a (non-measurable) function $f : \mathbb{R} \to \mathbb{R}$ with the following property: for any $g : \mathbb{R} \to \mathbb{R}$ such that

$$|g(x) - f(x)| < 1 \quad \text{for all } x \in \mathbb{R}, \tag{6.141}$$

g is not measurable.

Solution. Let V be a non-measurable set. We take

$$f = 2\chi_V. \tag{6.142}$$

Let $g : \mathbb{R} \to \mathbb{R}$ such that $|g(x) - f(x)| < 1$ for all $x \in \mathbb{R}$. Now, for $x \in V$

$$|g(x) - 2| < 1 \implies g(x) > 1. \tag{6.143}$$

For $x \notin V$,

$$|g(x) - 0| < 1 \implies g(x) < 1. \tag{6.144}$$

Now, $g^{-1}(1, \infty) = V$ is not measurable. Therefore, g can't be measurable. ■

Exercise 6.8. We can construct generalized Cantor sets as follows. If I is a bounded interval and $\alpha \in (0, 1)$, let us call the open interval with the same midpoint of I and length equal to α times the length of I the "open middle αth" of I. If $\{\alpha_j\}_{j\geq 1}$ is any sequence of numbers in $(0, 1)$, we can define a decreasing sequence $\{K_j\}_{j\geq 0}$ of closed sets as follows. Let $K_0 = [0, 1]$, and, for $j \geq 1$, K_j is obtained from K_{j-1} by removing the open middle α_j-th of each of the intervals that make up K_{j-1}. The resulting limiting set

$$K = \bigcap_{j=0}^{\infty} K_j$$

is our generalized Cantor set (sometimes also called "fat" Cantor set). The usual Cantor set arises when taking $\alpha_j = \frac{1}{3}$ for all j. Show that

(a) K is compact, has empty interior, and is totally disconnected (i.e. the only connected subsets of K are single points). Moreover, K has no isolated points.

(b) $m(K) = \prod_{j=1}^{\infty}(1 - \alpha_j)$.

(c) Show that $m(K) > 0$ if and only if $\sum_{j=1}^{\infty} \alpha_j < \infty$..

(d) Given $\beta \in (0, 1)$ exhibit a sequence $\{\alpha_j\}_{j\geq 1}$ such that $m(K) = \beta$.

Solution. Note that, K_n consists of 2^n intervals. Because at each step we are removing the middle α_j-th of each interval, so the number of intervals gets doubled at each step.

Also, we claim that the intervals of K_n has length $\ell_n = \prod_{j=1}^{n} \frac{1-\alpha_j}{2}$. We shall prove this by induction on n. It clearly holds for $n = 0$. Suppose it holds for some n. Then at the next step, each interval's middle α_{n+1}-th part gets removed, and we are left with the $1 - \alpha_{n+1}$, but that part gets divided into two parts. Therefore,

$$\ell_{n+1} = \frac{1 - \alpha_{n+1}}{2}\ \ell_n = \prod_{j=1}^{n+1} \frac{1 - \alpha_j}{2} \tag{6.145}$$

(a) K is bounded, and also closed (as it's the intersection of closed sets). Hence, it's compact.

Suppose $\operatorname{Int} K \neq \varnothing$. Then it contains an interval, say of size ε. Now, $\ell_n \leq 2^{-n} \to 0$, so we can find N such that $\ell_N < \varepsilon$. Since all the intervals in K_N has size $\ell_N < \varepsilon$, that means an interval of size ε can't be contained in K_N. Therefore, it can't be contained in K as it's the intersection of all K_n's.

Take two distinct points $x < y$. Suppose $\varepsilon = |x - y|$. As before, choose N such that $\ell_N < \varepsilon$. Now, $x, y \in K \subseteq K_N$. Since $\ell_N < \varepsilon$, x and y are in two different intervals in K_N. That means there was an interval (a, b) removed at one of the earlier stages. Then

$$K = \left[K \cap \left(-\infty, \frac{a+b}{2}\right)\right] \cup \left[K \cap \left(\frac{a+b}{2}, \infty\right)\right], \tag{6.146}$$

proving that x, y are separated. Therefore, K is totally separated, and in particular, totally disconnected.

Assume for the sake of contradiction that K has an isolated point x. Then there exists $\varepsilon > 0$ with the property that $B(x, \varepsilon)$ contain no point of $K - \{x\}$. As before, choose N such that $\ell_N < \varepsilon$. Since $x \in K \subseteq K_N$, so x is in one of the components of K_N, say $[a, b]$.

The length of the interval $[a, b]$ is ℓ_N, that means the distance from x to any of the points in this interval is at most ℓ_N, which is strictly smaller than ε. Therefore, $[a, b] \subseteq B(x, \varepsilon)$.

If $x \neq a$, $a \in B(x, \varepsilon) \cap (K - \{x\})$, since the endpoints are never removed. If $x = a$, $b \in B(x, \varepsilon) \cap (K - \{x\})$. Both cases, contradict with the assumption that $B(x, \varepsilon)$ contain no point of $K - \{x\}$.

(b) K_n is the union of closed intervals, so it is measurable. Therefore, K is also measurable as it is the intersection of countably many measurable sets. Now, K_n has 2^n disjoint intervals, each with length $\ell_n = \prod_{j=1}^{n} \frac{1-\alpha_j}{2}$. Therefore,

$$m(K_n) = 2^n \ell_n = \prod_{j=1}^{n} (1 - \alpha_j). \tag{6.147}$$

Now, K_n is a decreasing sequence of measurable sets, and K is their intersection, with $m(K_0) = 1 < \infty$. Therefore,

$$m(K) = \lim_{n\to\infty} m(K_n) = \lim_{n\to\infty} \prod_{j=1}^{n} (1 - \alpha_j) = \prod_{j=1}^{\infty} (1 - \alpha_j). \tag{6.148}$$

(c) We are gonna use the bound $1 - x \leq e^{-x}$ for $x \in (0, 1)$ and $1 - x \geq e^{-2x}$ for $x \in \left(0, \frac{1}{2}\right]$.

If $\sum_{j=1}^{\infty} \alpha_j$ diverges, then we have

$$\prod_{j=1}^{n} (1 - \alpha_j) \leq \prod_{j=1}^{n} \exp(-\alpha_n) = \exp\left(-\sum_{j=1}^{n} \alpha_n\right). \tag{6.149}$$

As $n \to \infty$, $\sum_{j=1}^{n} \alpha_n \to \infty$, so in that limit we have

$$m(K) = \lim_{n\to\infty} \prod_{j=1}^{n} (1 - \alpha_j) \leq \exp\left(-\lim_{n\to\infty} \sum_{j=1}^{n} \alpha_n\right) = 0. \tag{6.150}$$

Conversely, suppose $\sum_{j=1}^{\infty} \alpha_j$ converges. Then there exists some N such that $\alpha_n \leq \frac{1}{2}$ for $n \geq N$. Now,

$$\prod_{j=N}^{M} (1-\alpha_j) \geq \prod_{j=N}^{M} \exp(-2\alpha_n) = \exp\left(-2\sum_{j=N}^{M} \alpha_n\right). \tag{6.151}$$

As $M \to \infty$, the sum $\sum_{j \geq N} \alpha_n$ converges to a finite value, say s. Hence,

$$\prod_{j=N}^{\infty} (1-\alpha_j) \geq \exp\left(-2\sum_{j=N}^{\infty} \alpha_n\right) = e^{-2s} > 0. \tag{6.152}$$

Therefore, $\prod_{j=1}^{\infty} (1-\alpha_j) > 0$.

(d) Given $0 < \beta < 1$, we need to have

$$\beta = \prod_{j=1}^{\infty} (1-\alpha_j) \implies \log\beta = \sum_{j=1}^{\infty} \log(1-\alpha_j). \tag{6.153}$$

Therefore, we need to have that $\log(1-\alpha_j)$ sums to $\log\beta$. Indeed,

$$\log\beta = \log\beta \sum_{j=1}^{\infty} 2^{-j} = \sum_{j=1}^{\infty} \left[2^{-j}\log\beta\right], \tag{6.154}$$

so we may try $\log(1-\alpha_j) = 2^{-j}\log\beta$. Solving for α_j, we have

$$\alpha_j = 1 - \beta^{2^{-j}}. \tag{6.155}$$

This is clearly within $(0,1)$, and it satisfies $m(K) = \beta$.

■

Exercise 6.9. A measurable function is called *essentially bounded* if there exists a finite number $\beta \geq 0$ such that the set $\{x : |f(x)| > \beta\}$ has measure zero. Prove or disprove the following proposition:

> "Let $f : [0,1] \to \mathbb{R}$ be Lebesgue integrable. Then there exists $0 < a < b < 1$ such that f is essentially bounded in the interval (a,b)."

Solution. Let $q_1, q_2, q_3, \ldots$ be an enumeration of $\mathbb{Q} \cap [0,1]$. We define

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \cap [0,1], \\ \displaystyle\sum_{k=1}^{\infty} \frac{1}{2^k \sqrt{|x-q_k|}} & \text{otherwise.} \end{cases} \tag{6.156}$$

f is a non-negative function. So to show integrability, it suffices to show that $\int_0^1 f(x)\,\mathrm{d}x < \infty$.

Since the terms are all non-negative, by Proposition 2.8,

$$\int_0^1 f(x)\,\mathrm{d}x = \int_0^1 \sum_{k=1}^{\infty} \frac{1}{2^k\sqrt{|x-q_k|}}\,\mathrm{d}x = \sum_{k=1}^{\infty} \frac{1}{2^k}\int_0^1 \frac{1}{\sqrt{|x-q_k|}}\,\mathrm{d}x. \tag{6.157}$$

Let's now compute the integral:

$$\int_0^1 \frac{\mathrm{d}x}{\sqrt{|x-q_k|}} = \int_0^{q_k} \frac{\mathrm{d}x}{\sqrt{q_k-x}} + \int_{q_k}^1 \frac{\mathrm{d}x}{\sqrt{x-q_k}} = 2\sqrt{q_k} + 2\sqrt{1-q_k} \le 4. \tag{6.158}$$

Therefore,

$$\int_0^1 f(x)\,\mathrm{d}x \le \sum_{k=1}^{\infty} \frac{4}{2^k} = 4 < \infty. \tag{6.159}$$

Therefore, f is integrable. Now, assume for the sake of contradiction that f is essentially bounded on (a,b). Let's take a rational $q_m \in (0,1)$. Then for irrational x,

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{2^k\sqrt{|x-q_k|}} \ge \frac{1}{2^m\sqrt{|x-q_m|}} \tag{6.160}$$

Suppose f is essentially bounded by β on (a,b). Now, for irrational $x \in \left(q_m - \frac{1}{4^m\beta^2}, q_m + \frac{1}{4^m\beta^2}\right)$,

$$|x-q_m| < \frac{1}{4^m\beta^2} \implies \frac{1}{2^m\sqrt{|x-q_m|}} > \beta. \tag{6.161}$$

Therefore, $\{x : |f(x)| > \beta\}$ contains an interval (technically, all irrationals of an interval), so it has positive measure. Therefore, f can't be essentially bounded. ■

Exercise 6.10. Let $E \subset [0,1]$ be an open set, and let χ_E be its characteristic function. Prove or disprove: χ_E is Riemann integrable.

Solution. We shall use the following result:

> A bounded function $f : [0,1] \to \mathbb{R}$ is Riemann-integrable if and only if the set $\operatorname{disc} f$ of its discontinuities has measure 0.

Let q_n be an enumeration of the rational numbers in $(0,1)$. For every q_n, we can choose open intervals $(a_n, b_n) \subseteq (0,1)$ containing q_n such that

$$\sum_{n=1}^{\infty} (b_n - a_n) < 1. \tag{6.162}$$

This is because $\mathbb{Q} \cap (0,1)$ has measure 0, so we can cover it with intervals of arbitrarily small total length. Then we choose $E = \bigcup_{n=1}^{\infty} (a_n, b_n)$. Then E is an open set, with boundary

$$\partial E = \overline{E} - E = [0,1] - E. \tag{6.163}$$

Then $[0,1] = E \cup \partial E$, and

$$1 = m([0,1]) = m(E) + m(\partial E) \implies m(\partial E) = 1 - m(E) \geq 1 - \sum_{n=1}^{\infty} (b_n - a_n) > 0. \quad (6.164)$$

Therefore, the points of discontinuities of χ_E, i.e. the boundary of E, has positive measure. Hence, χ_E is **NOT** Riemann integrable. ■

Exercise 6.11. Proposition 3.8 says that for every measure space (Ω, Σ, μ), $f = 0$ a.e. if and only if $\int_A f \, \mathrm{d}\mu = 0$ for every measurable set A. Show that for if $f \in L^1(\mathbb{R})$, then

$$f = 0 \text{ a.e.} \iff \int_I f \, \mathrm{d}m = 0 \text{ for every interval } I \subset \mathbb{R}. \quad (6.165)$$

Solution. If $f = 0$ a.e., then it's trivially true that $\int_I f \, \mathrm{d}m = 0$. Conversely, suppose $\int_I f \, \mathrm{d}m = 0$ for every interval I. Assume for the sake of contradiction that f is not almost everywhere 0. Then $\{x \in \mathbb{R} : f(x) > 0\}$ or $\{x \in \mathbb{R} : f(x) < 0\}$ has positive measure. WLOG,

$$m(S) := m(\{x \in \mathbb{R} : f(x) > 0\}) > 0. \quad (6.166)$$

Claim 6.8: If $U \subseteq \mathbb{R}$ is an open set, then $\int_U f \, \mathrm{d}m = 0$.

Proof. Every open set is a disjoint union of intervals, so if $U = \bigcup_{n=1}^{\infty} (a_n, b_n)$, then

$$\int_U f \, \mathrm{d}m = \sum_{n=1}^{\infty} \int_{(a_n, b_n)} f \, \mathrm{d}m. = 0, \quad (6.167)$$

since $\int_{(a_n,b_n)} f \, \mathrm{d}m = 0$, by hypothesis. □

We write

$$S = \bigcup_{M=1}^{\infty} (S \cap (-M, M)). \quad (6.168)$$

Since S has positive measure, then there exists some M such that

$$m(S \cap (-M, M)) > 0, \quad (6.169)$$

because otherwise, S would be a countable union of nullsets, which is also a nullset. Choose

$$S' = S \cap (-M, M). \quad (6.170)$$

Then there exists some closed set $F \subseteq S'$ such that

$$m(S' - F) \leq \varepsilon := \frac{m(S')}{2}. \quad (6.171)$$

Then

$$\frac{m(S')}{2} \geq m\left(S' - F\right) = m\left(S'\right) - m\left(F\right) \implies m(F) \geq \frac{m(S')}{2} > 0. \tag{6.172}$$

Now, $F \subseteq S' \subseteq S$, so that $f > 0$ on F, and $m(F) > 0$. By Claim 3.1,

$$\int_F f \, \mathrm{d}m > 0. \tag{6.173}$$

Now, $F \subseteq S' \subseteq (-M, M)$, so

$$0 = \int_{(-M,M)} f \, \mathrm{d}m = \int_{(-M,M)-F} f \, \mathrm{d}m + \int_F f \, \mathrm{d}m \implies \int_F f \, \mathrm{d}m = 0, \tag{6.174}$$

since $(-M, M) - F$ is an open set. Therefore, (6.173) and (6.174) contradict each other! ■

§7. Extra material: General outer measure

We show other important ideas to construct general measures starting from an outer measure. In particular, the discussion here leads us to an alternative way of constructing the Lebesgue measure.

§7.i. Outer measures

Definition 7.1 (Outer measure)**.** An **outer measure** μ^* on a set Ω is a function $\mu^* : \mathcal{P}(\Omega) \to [0, \infty]$ satisfying the following properties:

(i) $\mu^*(\varnothing) = 0$;

(ii) $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$;

(iii) μ^* is countably sub-additive, i.e.

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n). \tag{7.1}$$

We now describe a particularly useful way to generate outer measures. You may start with a collection of "elementary" sets $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ that you "know" how to measure (e.g. balls or rectangles in $\mathbb{R}^d$) and then approximate generic sets from the outside by countable unions of such elementary sets.

Proposition 7.1 (Generating outer measures)

Let $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ and $\rho : \mathcal{E} \to [0, \infty]$ be such that $\varnothing \in \mathcal{E}$, $\Omega \in \mathcal{E}$ and $\rho(\varnothing) = 0$. For any $A \subseteq \Omega$ define

$$\mu^*(A) = \inf\left\{\sum_{n=1}^{\infty} \rho(E_n) : E_n \in \mathcal{E} \text{ and } A \subseteq \bigcup_{n=1}^{\infty} E_n\right\}. \tag{7.2}$$

Then μ^* is an outer measure.

Proof. From the definition it is clear that $\mu^*(\varnothing) = 0$ and that $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$. To prove the countable sub-additivity we argue as follows. Suppose $\{A_j\}_{j=1}^{\infty} \subseteq \mathcal{P}(\Omega)$ and let $\varepsilon > 0$ be given. By definition, for each j there exists a collection $\{E_j^n\}_{n=1}^{\infty} \subseteq \mathcal{E}$ such that $A_j \subseteq \bigcup_{n=1}^{\infty} E_j^n$ and

$$\sum_{n=1}^{\infty} \rho(E_j^n) \leq \mu^*(A_j) + \frac{\varepsilon}{2^j}. \tag{7.3}$$

If $A = \bigcup_{j=1}^{\infty} A_j$ we then have $A \subseteq \bigcup_{j,n=1}^{\infty} E_j^n$ and hence

$$\mu^*(A) \leq \sum_{j,n=1}^{\infty} \rho(E_j^n) \leq \sum_{j=1}^{\infty} \mu^*(A_j) + \varepsilon. \tag{7.4}$$

Since $\varepsilon > 0$ is arbitrary, we are done. ∎

Remark 7.1. Note that, at this point, we are not claiming that $\mu^*(E) = \rho(E)$ for a set $E \in \mathcal{E}$.

§7.ii. Carathéodory's theorem

There is a beautiful and fundamental result that leads us from outer measures to measures as we now describe.

Definition 7.2. If μ^* is an outer measure on Ω, a set $A \subseteq \Omega$ will be called μ^*-measurable if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{7.5}$$

for every $E \subseteq \Omega$.

Remark 7.2. From the definition of outer measure, note that we always have

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c), \tag{7.6}$$

so it is the other inequality

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{7.7}$$

that needs to be verified in order for a set A to be μ^*-measurable.

Theorem 7.2 (Carathéodory's theorem)

If μ^* is an outer measure on Ω, the collection Σ^* of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to Σ^* is a complete measure.

Proof. Let us divide this proof into its core steps.

Step 1: Σ^* is closed under complements. This follows from the fact that definition (7.5) is symmetric with respect to A and A^c.

Step 2: Σ^* is closed under finite unions and μ^* is finitely additive. Let $A, B \in \Sigma^*$ and $E \subseteq \Omega$. Then, from the definition

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).\end{aligned} \tag{7.8}$$

Noting that $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, and using the sub-additivity, we have

$$\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \geq \mu^*(E \cap (A \cup B)). \tag{7.9}$$

From (7.8) and (7.9) we get

$$\mu^*(E) \geq \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c), \tag{7.10}$$

and it follows that $A \cup B \in \Sigma^*$. From this we get that Σ^* is closed under finite unions.

If $A, B \in \Sigma^*$ and $A \cap B = \varnothing$, we have

$$\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B). \tag{7.11}$$

Step 3: Σ^* is closed under countable unions and μ^* is countably additive. Now let $\{A_j\}_{j=1}^{\infty} \subseteq \Sigma^*$; we want to show that $\bigcup_{j=1}^{\infty} A_j \in \Sigma^*$. Note that it suffices to prove this claim for A_j's that are disjoint, since in the generic case we could consider

$$\widetilde{A}_1 = A_1, \text{ and } \widetilde{A}_j = A_j - (A_1 \cup \cdots \cup A_{j-1}). \tag{7.12}$$

Now, $\widetilde{A}_j$'s that are disjoint and

$$\bigcup_{j=1}^{n} \widetilde{A}_j = \bigcup_{j=1}^{n} A_j \tag{7.13}$$

for each n, and

$$\bigcup_{j=1}^{\infty} \widetilde{A}_j = \bigcup_{j=1}^{\infty} A_j. \tag{7.14}$$

So let us assume then that our original A_j's are disjoint. Let

$$B_n = \bigcup_{j=1}^{n} A_j, \quad B = \bigcup_{j=1}^{\infty} A_j. \tag{7.15}$$

(set $B_0 = \varnothing$). Then for any $E \subseteq \Omega$, we have

$$\begin{aligned} \mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}). \end{aligned} \tag{7.16}$$

From (7.16) we can easily show by induction that

$$\mu^*(E \cap B_n) = \sum_{j=1}^{n} \mu^*(E \cap A_j). \tag{7.17}$$

Therefore, since $B_n \subseteq B$, $B_n^c \supseteq B^c$, and using monotonicity,

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geq \sum_{j=1}^{n} \mu^*(E \cap A_j) + \mu^*(E \cap B^c). \tag{7.18}$$

Letting $n \to \infty$ in (7.18) we get

$$\begin{aligned}\mu^*(E) &\geq \sum_{j=1}^{\infty} \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \\ &\geq \mu^*\left(\bigcup_{j=1}^{\infty}(E \cap A_j)\right) + \mu^*(E \cap B^c) \\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c), \end{aligned} \tag{7.19}$$

and we conclude that $B \in \Sigma^*$. Moreover, all the inequalities in (7.19) are in fact equalities. Taking $E = B$ in (7.19) we obtain

$$\mu^*(B) = \sum_{j=1}^{\infty} \mu^*(A_j) \tag{7.20}$$

which is the desired countable additivity.

Step 4: The measure μ^* is complete. If $\mu^*(A) = 0$, for any $E \subseteq \Omega$ we have

$$\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leq \mu^*(E). \tag{7.21}$$

Hence $A \in \Sigma^*$. This shows that μ^* restricted to the σ-algebra Σ^* is indeed a complete measure. This concludes the proof. ■

At this point we cannot yet say too much about the σ-algebra Σ^* of μ^*-measurable sets. It could be the case, for example, that we have $\Sigma^* = \{\varnothing, \Omega\}$. We need a little bit more structure in order to apply Carathéodory's theorem in a way to generate concrete σ-algebras and measures.

§7.iii. Pre-measures

Definition 7.3 (Algebra)**.** We say that a non-empty collection $\mathcal{A}$ of subsets of Ω is an **algebra** if it is closed under complements and finite unions, that is

(i) If A belongs to $\mathcal{A}$ then its complement $A^c := \Omega - A$ also belongs to $\mathcal{A}$;

(ii) If $\{A_n\}_{n=1}^N \subseteq \mathcal{A}$ then $\bigcup_{n=1}^N A_n \in \mathcal{A}$.

Definition 7.4 (Pre-measure)**.** Let $\mathcal{A}$ be an algebra. A function $\mu_0 : \mathcal{A} \to [0, \infty]$ is called a **pre-measure** if the following conditions are verified:

(i) $\mu_0(\varnothing) = 0$;

(ii) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of disjoint sets in $\mathcal{A}$ such that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then

$$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n). \tag{7.22}$$

In particular, the pre-measure is finitely additive since we can always take $A_n = \varnothing$ for $n > N$. The notions of finite and σ-finite pre-measures are the same as we have discussed for measures.

If μ_0 is a pre-measure on the algebra $\mathcal{A}$, we have seen in Proposition 7.1 that

$$\mu^*(E) = \inf\left\{\sum_{n=1}^{\infty}\mu_0(A_n) : A_n \in \mathcal{A} \text{ and } E \subseteq \bigcup_{n=1}^{\infty} A_n\right\} \tag{7.23}$$

defines an outer measure on Ω. In this setup we can prove the following.

Proposition 7.3

If μ_0 is a pre-measure on the algebra $\mathcal{A}$, and μ^* is the outer measure defined by (7.23), then

(i) $\mu^*|_{\mathcal{A}} = \mu_0$.

(ii) Every set in $\mathcal{A}$ is μ^*-measurable.

Proof. (i) Let $E \in \mathcal{A}$. Clearly, since E covers itself (i.e. we may take $A_1 = E$ and $A_j = \varnothing$ for $j \geq 2$), we have

$$\mu^*(E) \leq \mu_0(E). \tag{7.24}$$

If $E \subseteq \bigcup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{A}$, we may consider

$$\widetilde{A}_1 = E \cap A_1, \text{ and } \widetilde{A}_j = E \cap \left(A_j - \bigcup_{i=1}^{j-1} A_i\right). \tag{7.25}$$

Then $\widetilde{A}_j$'s are disjoint members of $\mathcal{A}$ whose union is E and we get

$$\mu_0(E) = \sum_{n=1}^{\infty}\mu_0\left(\widetilde{A}_j\right) \leq \sum_{n=1}^{\infty}\mu_0(A_j). \tag{7.26}$$

Hence it follows that $\mu_0(E) \leq \mu^*(E)$. Therefore, $\mu_0(E) = \mu^*(E)$.

(ii) If $A \in \mathcal{A}$ and $E \subseteq \Omega$, and $\varepsilon > 0$ is given, there is a sequence $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{A}$ such that

$$E \subseteq \bigcup_{n=1}^{\infty} A_n, \text{ and } \sum_{n=1}^{\infty}\mu_0(A_n) \leq \mu^*(E) + \varepsilon. \tag{7.27}$$

Since μ_0 is additive on $\mathcal{A}$, for each A_n,

$$\mu_0(A_n) = \mu_0(A_n \cap A) + \mu_0(A_n \cap A^c). \tag{7.28}$$

As a result,

$$\mu^*(E) + \varepsilon \geq \sum_{n=1}^{\infty}\mu_0(A_n) = \sum_{n=1}^{\infty}\mu_0(A_n \cap A) + \sum_{n=1}^{\infty}\mu_0(A_n \cap A^c). \tag{7.29}$$

Now, since $\{A_n \cap A\}_{n\geq 1} \subseteq \mathcal{A}$ covers $E \cap A$,

$$\mu^*(E \cap A) \leq \sum_{n=1}^{\infty} \mu_0(A_n \cap A). \tag{7.30}$$

Similarly,

$$\mu^*(E \cap A^c) \leq \sum_{n=1}^{\infty} \mu_0(A_n \cap A^c). \tag{7.31}$$

Combining all these, we find

$$\mu^*(E) + \varepsilon \geq \sum_{n=1}^{\infty} \mu_0(A_n \cap A) + \sum_{n=1}^{\infty} \mu_0(A_n \cap A^c) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c). \tag{7.32}$$

Since $\varepsilon > 0$ is arbitrary, $\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c)$. This shows that A is μ^*-measurable.

■

With this in mind, the construction given by Carathéodory's theorem applied to a pre-measure μ_0 will give us a powerful result.

Theorem 7.4

Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be an algebra, μ_0 a pre-measure on $\mathcal{A}$, and Σ be the σ-algebra generated by $\mathcal{A}$. There exists a measure μ on Σ whose restriction to $\mathcal{A}$ is μ_0, namely $\mu = \mu^*|_\Sigma$, with μ^* given by (7.23).

If ν is another measure on Σ that extends μ_0 then $\nu(E) \leq \mu(E)$ for all $E \in \Sigma$ with equality if $\mu(E) < \infty$. If μ_0 is σ-finite, then μ is the unique extension of μ_0 to a measure on Σ.

Proof. The first assertion follows immediately from Carathéodory's theorem and Proposition 7.3, since the σ-algebra Σ^* of μ^*-measurable sets contains $\mathcal{A}$ and hence the σ-algebra Σ generated by $\mathcal{A}$.

For the second assertion, if $E \in \Sigma$ and $E \subseteq \bigcup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{A}$, then

$$\nu(E) \leq \sum_{n=1}^{\infty} \nu(A_n) = \sum_{n=1}^{\infty} \mu_0(A_n). \tag{7.33}$$

Since $\mu(E) = \inf \sum_{n=1}^{\infty} \mu_0(A_n)$, we can take infimum on the RHS and we get

$$\nu(E) \leq \inf_{\substack{\{A_n\}_n \subseteq \mathcal{A}, \\ E \subseteq \bigcup_n A_n}} \sum_{n=1}^{\infty} \mu_0(A_n) = \mu(E). \tag{7.34}$$

If we set $A = \bigcup_{n=1}^{\infty} A_n$ we have

$$\nu(A) = \lim_{N\to\infty} \nu\left(\bigcup_{n=1}^{N} A_n\right) = \lim_{N\to\infty} \mu\left(\bigcup_{n=1}^{N} A_n\right) = \mu(A). \tag{7.35}$$

If $\mu(E) < \infty$ and $\varepsilon > 0$ is given, we can choose A above such that

$$\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu(E) + \varepsilon. \tag{7.36}$$

Since $\mu(A) = \mu(A - E) + \mu(E)$ we get $\mu(A - E) \leq \varepsilon$ and

$$\mu(E) \leq \mu(A) = \nu(A) = \nu(E) + \nu(A - E) \leq \nu(E) + \mu(A - E) \leq \nu(E) + \varepsilon. \tag{7.37}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\nu(E) = \mu(E)$.

Finally, assume that $\Omega = \bigcup_{n=1}^{\infty} A_n$ with $\mu_0(A_n) < \infty$. We may also assume that all the A_n's are disjoint. Then, for any $E \in \Sigma$, we have

$$\mu(E) = \sum_{n=1}^{\infty} \mu(E \cap A_n) = \sum_{n=1}^{\infty} \nu(E \cap A_n) = \nu(E). \tag{7.38}$$

This concludes the proof. ■

Remark 7.3. Observe that μ_0 can actually be extended to a measure μ^* on the σ-algebra Σ^*.

§7.iv. Exercises

Exercise 7.1. Let $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be an algebra, $\mathcal{A}_\sigma$ be the collection of countable unions of sets in $\mathcal{A}$ and $\mathcal{A}_{\sigma\delta}$ be the collection of countable intersections of sets in $\mathcal{A}_\sigma$. Let μ_0 be a pre-measure on $\mathcal{A}$ and μ^* be the induced outer measure. Show that:

(i) For any $E \subseteq \Omega$ and $\varepsilon > 0$ there exists $A \in \mathcal{A}_\sigma$ with $E \subseteq A$ and $\mu^*(A) \leq \mu^*(E) + \varepsilon$.

(ii) If $\mu^*(E) < \infty$ then E is μ^*-measurable if and only if there exists $B \in \mathcal{A}_{\sigma\delta}$ with $E \subseteq B$ and $\mu^*(B - E) = 0$.

(iii) If μ_0 is σ-finite, show that the condition $\mu^*(E) < \infty$ in (ii) is not necessary.

Solution. (i) By definition of outer measure as an infimum, there exists $\{A_n\}_n \subseteq \mathcal{A}$ such that

$$\sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu^*(E) + \varepsilon. \tag{7.39}$$

WLOG, we can assume A_n's are pairwise disjoint. If not, we can do a similar construction as in (7.12). We claim that $A = \bigcup_{n=1}^{\infty} A_n$ is the desired set. Notice that $A_n \in \mathcal{A}$ implies

they are μ^*-measurable. Then so is A. Then by continuity,

$$\mu^*(A) = \mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{N\to\infty} \mu^*\left(\bigcup_{n=1}^{N} A_n\right) = \lim_{N\to\infty} \mu_0\left(\bigcup_{n=1}^{N} A_n\right). \tag{7.40}$$

Since $\mathcal{A}$ is closed under finite union, and μ_0 is finitely additive,

$$\mu^*(A) = \lim_{N\to\infty} \sum_{n=1}^{N} \mu_0(A_n) = \sum_{n=1}^{\infty} \mu_0(A_n) \leq \mu^*(E) + \varepsilon. \tag{7.41}$$

(ii) Suppose E is measurable. By part (i), there exists A_n with $E \subseteq A_n$ and

$$\mu^*(A_n) \leq \mu^*(E) + \frac{1}{n}. \tag{7.42}$$

Consider $B = \bigcap_{n=1}^{\infty} A_n$. This contains E. Then

$$\mu^*(E) \leq \mu^*(B) \leq \mu^*(A_n) \leq \mu^*(E) + \frac{1}{n}, \tag{7.43}$$

which is true for every n. Therefore, $\mu^*(B) = \mu^*(E)$. Since E is measurable,

$$\mu^*(B) = \mu^*(B \cap E) + \mu^*(B \cap E^c) \implies \mu^*(B - E) = \mu^*(B) - \mu^*(E) = 0. \tag{7.44}$$

Conversely, suppose there exists $B \in \mathcal{A}_{\sigma\delta}$ such that $E \subseteq B$ and $\mu^*(B - E) = 0$. Since μ^* is a complete measure, $B - E$ is measurable. Then so is

$$B - (B - E) = E. \tag{7.45}$$

(iii) The if direction doesn't require $\mu^*(E)$ to be finite. So we'll prove the other direction. Suppose E is measurable.

Since μ_0 is σ-finite, there's a sequence of sets $X_1, X_2, \ldots$, each with finite μ_0-measure, and $\Omega = \bigcup_{n=1}^{\infty} X_n$. WLOG, they are disjoint. $X_n \in \mathcal{A}$, so they are measurable. Therefore, $E \cap X_n$ is also measurable.

Given $\varepsilon > 0$, we can find $B_n \in \mathcal{A}_\sigma$ such that $B_n \supseteq E \cap X_n$, and

$$\mu^*(B_n) \leq \mu^*(E \cap X_n) + \frac{\varepsilon}{2^n}. \tag{7.46}$$

Since $(E \cap X_n)$ has finite measure,

$$\mu^*(B_n - (E \cap X_n)) = \mu^*(B_n) - \mu^*(E \cap X_n) \leq \frac{\varepsilon}{2^n}. \tag{7.47}$$

By subadditivity,

$$\mu^*\left(\bigcup_{n=1}^{\infty} (B_n - (E \cap X_n))\right) \leq \sum_{n=1}^{\infty} \mu^*(B_n - (E \cap X_n)) \leq \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon. \tag{7.48}$$

Now,

$$\left(\bigcup_{n=1}^{\infty} B_n\right) - E = \left(\bigcup_{n=1}^{\infty} B_n\right) - \left(\bigcup_{n=1}^{\infty} (E \cap X_n)\right) \subseteq \bigcup_{n=1}^{\infty} (B_n - (E \cap X_n)). \tag{7.49}$$

Therefore, $\mu^*\left(\left(\bigcup_{n=1}^{\infty} B_n\right) - E\right) \leq \varepsilon$, where $E \subseteq \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}_{\sigma\sigma} = \mathcal{A}_\sigma$. Therefore, for each n, we can find $A_n \in \mathcal{A}_\sigma$ such that $E \subseteq A_n$ and

$$\mu^*(A_n - E) \leq \frac{1}{n}. \tag{7.50}$$

Then we take

$$B = \bigcap_{n=1}^{\infty} A_n \in A_{\sigma\delta}. \tag{7.51}$$

Clearly, $E \subseteq B$. Now,

$$\mu^*(B - E) = \mu^*\left(\left(\bigcap_{n=1}^{\infty} A_n\right) - E\right) \leq \mu^*(A_n - E) \leq \frac{1}{n}, \tag{7.52}$$

which is true for any n. Therefore, $\mu^*(B - E) = 0$.

■

Exercise 7.2. Let μ^* be an outer measure on Ω, Σ^* be the σ-algebra of μ^*-measurable sets, $\mu = \mu^*|_{\Sigma^*}$, and μ^+ be the outer measure induced by μ as in (7.23) (with μ and Σ^* replacing μ_0 and $\mathcal{A}$ in that definition). Show that

(i) If $E \subseteq \Omega$, we have $\mu^*(E) \leq \mu^+(E)$, with equality if and only if there exists $A \in \Sigma^*$ with $E \subseteq A$ and $\mu^*(A) = \mu^*(E)$.

(ii) If μ^* is induced by a pre-measure then $\mu^* = \mu^+$.

(iii) If $\Omega = \{0, 1\}$, construct an example such that $\mu^* \neq \mu^+$.

Solution. (i) Let $\{A_n\}_n \subseteq \Sigma^*$ such that $E \subseteq \bigcup_{n=1}^{\infty} A_n$. By monotonicity and subadditivity,

$$\mu^*(E) \leq \mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n) = \sum_{n=1}^{\infty} \mu(A_n), \tag{7.53}$$

since $A_n \in \Sigma^*$ and $\mu = \mu^*|_{\Sigma^*}$. Now, taking infimum on the RHS over all $\{A_n\}_n \subseteq \Sigma^*$ such that $E \subseteq \bigcup_{n=1}^{\infty} A_n$, we get

$$\mu^*(E) \leq \inf_{\substack{A_n \in \Sigma^*, \\ E \subseteq \bigcup_n A_n}} \sum_{n=1}^{\infty} \mu(A_n) = \mu^+(E). \tag{7.54}$$

If there exists $A \in \Sigma^*$ with $E \subseteq A$ and $\mu^*(A) = \mu^*(E)$, then A covers E, so

$$\mu^+(E) \leq \mu(A) = \mu^*(A) = \mu^*(E), \tag{7.55}$$

proving that $\mu^*(E) = \mu^+(E)$.

Now, conversely, suppose $\mu^*(E) = \mu^+(E)$. Given n, by the definition of μ^+, there exists $\{B_{n,m}\}_m \subseteq \Sigma^*$ such that $E \subseteq \bigcup_{m=1}^{\infty} B_{n,m} =: A_n$ and

$$\sum_{m=1}^{\infty} \mu\left(B_{n,m}\right) \leq \mu^+(E) + \frac{1}{n} = \mu^*(E) + \frac{1}{n}. \tag{7.56}$$

The LHS is at least $\mu\left(A_n\right)$, by subadditivity. Now we take $A = \bigcap_{n=1}^{\infty} A_n \in \Sigma^*$. Then A contains E, so by monotonicity, $\mu^*(E) \leq \mu^*\left(A\right)$. Now,

$$\mu^*\left(A\right) \leq \mu^*\left(A_n\right) \leq \sum_{m=1}^{\infty} \mu\left(B_{n,m}\right) \leq \mu^*(E) + \frac{1}{n}, \tag{7.57}$$

which is true for every n, proving that $\mu^*\left(A\right) = \mu^*\left(E\right)$. Therefore, $\mu^*\left(A\right) = \mu^*\left(E\right)$.

(ii) Suppose μ^* is induced by a pre-measure μ_0 on $\mathcal{A}$. Take $E \subseteq \Omega$. By Exercise 7.1(i), we can find $A_n \in \mathcal{A}_\sigma \subseteq \Sigma^*$ with $E \subseteq A_n$ and

$$\mu^*\left(A_n\right) \leq \mu^*\left(E\right) + \frac{1}{n}. \tag{7.58}$$

Now we take $A = \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}_{\sigma\delta} \subseteq \Sigma^*$. Then A contains E, so by monotonicity, $\mu^*(E) \leq \mu^*\left(A\right)$. Now,

$$\mu^*\left(A\right) \leq \mu^*\left(A_n\right) \leq \sum_{m=1}^{\infty} \mu\left(B_{n,m}\right) \leq \mu^*(E) + \frac{1}{n}, \tag{7.59}$$

which is true for every n, proving that $\mu^*\left(A\right) = \mu^*\left(E\right)$. Therefore, $\mu^*\left(A\right) = \mu^*\left(E\right)$. Then by part (ii), $\mu^*\left(E\right) = \mu^+(E)$.

(iii) We define the outer measure μ^* on $\mathcal{P}\left(\Omega\right)$ as follows:

$$\mu^*\left(\varnothing\right) = 0, \quad \mu^*\left(\{0\}\right) = 2, \quad \mu^*\left(\{1\}\right) = 3, \quad \mu^*\left(\{0,1\}\right) = 4. \tag{7.60}$$

This is an outer measure, as it satisfies monotonicity, and subadditivity. Clearly, $\varnothing, \Omega$ are μ^*-measurable. We claim that these are the only measurable sets. Indeed, if $\{0\}$ was measurable, then we needed to have

$$\mu^*\left(\{0,1\}\right) = \mu^*\left(\{0,1\} \cap \{0\}\right) + \mu^*\left(\{0,1\} \cap \{0\}^c\right), \tag{7.61}$$

but which is **NOT** true as $4 \neq 2 + 3$. Therefore, $\Sigma^* = \{\varnothing, \Omega\}$.

Finally, to compute $\mu^+\left(\{0\}\right)$, there's only one set in Σ^* that covers $\{0\}$, namely $\{0,1\}$. As a result, $\mu^+\left(\{0\}\right) = 4 \neq \mu^*\left(\{0\}\right)$.

■

Exercise 7.3. Let (Ω, Σ, μ) be a measure space, μ^* be the outer measure induced by μ as in (7.23), Σ^* be the σ-algebra of μ^*-measurable sets and $\overline{\mu} = \mu^*|_{\Sigma^*}$. Show that if μ is σ-finite then $\overline{\mu}$ is the completion of μ.

Solution. $\overline{\mu}$ is a complete measure on Σ^*, by Carathéodory's theorem. Let $\widehat{\mu}$ be the completion on μ, whose domain is

$$\overline{\Sigma} = \{E \cup F : E \in \Sigma,\ F \subseteq N \text{ for some } N \in \Sigma \text{ with } \mu(N) = 0\}. \tag{7.62}$$

$\widehat{\mu}$ is the unique completion of μ on $\overline{\Sigma}$. In order to show that $\overline{\mu} = \widehat{\mu}$, since both are complete measures and both of them extend μ, it suffices to show that their domains are equal. So we'll show $\Sigma^* = \overline{\Sigma}$.

Take $E \cup F \in \overline{\Sigma}$, where $E \in \Sigma$ and $F \subseteq N$ for some $N \in \Sigma$ with $\mu(N) = 0$. Then $E \cup N \in \Sigma$. Now,

$$(E \cup N) - (E \cup F) \subseteq N - F. \tag{7.63}$$

As a result,

$$\mu^*((E \cup N) - (E \cup F)) \leq \mu^*(N - F) \leq \mu^*(N) = 0, \tag{7.64}$$

proving that $\mu^*((E \cup N) - (E \cup F)) = 0$. Now, $E \cup N \in \Sigma = \Sigma_{\sigma\delta}$. Since μ is σ-finite, by Exercise 7.1(iii), $E \cup F$ is μ^*-measurable, i.e. $E \cup F \in \Sigma^*$.

Now, suppose $X \in \Sigma^*$. By Exercise 7.1(iii), we can find $B \in \Sigma_{\sigma\delta} = \Sigma$ such that

$$X \subseteq B, \text{ and } \mu^*(B - X) = 0. \tag{7.65}$$

This proves that $B - X$ is μ^*-measurable, since it's a complete measure. Then again by Exercise 7.1(iii), we can find $E \in \Sigma_{\sigma\delta} = \Sigma$ such that

$$B - X \subseteq E, \text{ and } \mu^*(E - (B - X)) = 0. \tag{7.66}$$

Now, by subadditivity,

$$\mu(E) = \mu^*(E) \leq \mu^*(B - X) + \mu^*(E - (B - X)) = 0. \tag{7.67}$$

Now, we write

$$X^c = \Omega - X = (\Omega - B) \cup (B - X). \tag{7.68}$$

$B \in \Sigma$, so $\Omega - B \in \Sigma$. Furthermore, $B - X$ is contained in a μ-measure zero set E. Therefore, $X^c \in \overline{\Sigma}$. Since $\overline{\Sigma}$ is a σ-algebra, $X \in \overline{\Sigma}$. ■

§7.v. Borel measures on $\mathbb{R}$

Let $\mathcal{B}_{\mathbb{R}}$ be the Borel σ-algebra on $\mathbb{R}$. A little bit of motivation for what we are going to do: suppose that μ is a measure on $\mathcal{B}_{\mathbb{R}}$ we can consider the distribution function $F : \mathbb{R} \to \mathbb{R}$

$$F(x) = \mu((-\infty, x]). \tag{7.69}$$

One can see that this function is non-decreasing and right-continuous, that is $\lim_{y \to x^+} F(y) = F(x)$. Indeed, if $x_n \to x$ decreasingly, then

$$(-\infty, x] = \bigcap_{n=1}^{\infty} (-\infty, x_n]\,. \tag{7.70}$$

By continuity of measure,

$$F(x) = \mu((-\infty, x]) = \lim_{n\to\infty} \mu((-\infty, x_n]) = \lim_{n\to\infty} F(x_n)\,. \tag{7.71}$$

So $F : \mathbb{R} \to \mathbb{R}$, defined as $F(x) = \mu((-\infty, x])$ is non-decreasing and right-continuous. We want to revert this process and construct a measure μ starting from a non-decreasing and right-continuous function $F : \mathbb{R} \to \mathbb{R}$.

We will consider the collection of half-open intervals to the left, called here simply h-intervals. These will be intervals of the following form: $(a, b]$, (a, ∞) or $\varnothing$, where $-\infty \le a < b < \infty$.

Lemma 7.5

The collection $\mathcal{A}$ of finite disjoint unions of h-intervals is an algebra. Furthermore, the σ-algebra generated by $\mathcal{A}$ is $\mathcal{B}_{\mathbb{R}}$.

Proof. Let $\mathcal{I}_h$ be the collection of "basic" h-intervals,

$$\mathcal{I}_h = \{(a, b], (a, \infty), \varnothing : -\infty \le a < b < \infty\}, \tag{7.72}$$

and $\mathcal{A}$ is the collection of finite disjoint unions of elements of $\mathcal{I}_h$. Clearly, $\varnothing, \mathbb{R} \in \mathcal{A}$, as they are in $\mathcal{I}_h$.

$\mathcal{A}$ is closed under finite intersection. Let $I, J \in \mathcal{I}_h$. Then we have the following cases:

- $I = (a, b], J = (c, d]$. Then $I \cap J = (\max\{a, c\}, \min\{b, d\}]$.
- $I = (a, \infty), J = (c, \infty)$. Then $I \cap J = (\max\{a, c\}, \infty)$.
- $I = (a, b], J = J = (c, \infty)$. Then $I \cap J = (\max\{a, c\}, b]$ or $\varnothing$.

Therefore, the intersection of basic h-intervals is again a basic h-interval. Now, pick $A = \bigcup_{i=1}^{n} I_i, B = \bigcup_{j=1}^{m} J_j \in \mathcal{A}$, where I_i's are pairwise disjoint h-intervals and J_j's are pairwise

disjoint h-intervals.

$$A \cap B = \left(\bigcup_{i=1}^{n} I_i \right) \cap \left(\bigcup_{j=1}^{m} J_j \right) = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} (I_i \cap J_j) \,. \tag{7.73}$$

$I_i \cap J_j$ is again a basic h-interval, and they are all pairwise disjoint. Therefore, $A \cap B \in \mathcal{A}$.

$\mathcal{A}$ is closed under complement. Let $I \in \mathcal{I}_h$. Then we have the following cases:

- $I = (a, b]$. Then $I^c = (-\infty, a] \cup (b, \infty) \in \mathcal{A}$.
- $I = (a, \infty)$. Then $I^c = (-\infty, a] \in \mathcal{A}$.

Therefore, the complement of basic h-intervals is in $\mathcal{A}$. Now, pick $A = \bigcup_{i=1}^{n} I_i \in \mathcal{A}$, where I_i's are pairwise disjoint h-intervals.

$$A^c = \left(\bigcup_{i=1}^{n} I_i \right)^c = \bigcap_{i=1}^{n} I_i^c. \tag{7.74}$$

$I_i^c \in \mathcal{A}$ and $\mathcal{A}$ is closed under finite intersection. Therefore, $A^c \in \mathcal{A}$. Hence, $\mathcal{A}$ is an algebra.

$\mathcal{A}$ generates $\mathcal{B}_{\mathbb{R}}$. Clearly $\sigma(\mathcal{A}) \subseteq \mathcal{B}_{\mathbb{R}}$. Now, we show the reverese inclusion. Note that any open interval (a, b) can be written as

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n} \right] \in \sigma(\mathcal{A}). \tag{7.75}$$

Since any open set is the countable union of such intervals, we can say that any open set is in $\sigma(\mathcal{A})$. Hence, $\mathcal{B}_{\mathbb{R}} \subseteq \sigma(\mathcal{A})$. ■

Lemma 7.6

Let $F : \mathbb{R} \to \mathbb{R}$ be a non-decreasing and right-continuous function. If $(a_j, b_j]$ are disjoint h-intervals $(j = 1, 2, 3, \ldots, n)$ let

$$\mu_0 \left(\bigcup_{j=1}^{n} (a_j, b_j] \right) = \sum_{j=1}^{n} (F(b_j) - F(a_j)) \tag{7.76}$$

and let $\mu_0(\varnothing) = 0$. Then μ_0 is well-defined and defines a pre-measure on $\mathcal{A}$.

Proof. First we must check that μ_0 is well defined, since elements of $\mathcal{A}$ can be represented in more than one way as disjoint unions of h-intervals. If $\{(a_j, b_j]\}_1^n$ are disjoint and $\bigcup_1^n (a_j, b_j] = (a, b]$, then, after perhaps relabeling the index j, we must have

$$a = a_1 < b_1 = a_2 < b_2 = \cdots < b_n = b, \tag{7.77}$$

so

$$\sum_{j=1}^{n} \left[F\left(b_j\right) - F\left(a_j\right)\right] = F\left(b\right) - F\left(a\right). \tag{7.78}$$

More generally, if $\{I_i\}_1^n$ and $\{J_j\}_1^m$ are finite sequences of disjoint h-intervals such that $\bigcup_1^n I_i = \bigcup_1^m J_j$, this reasoning shows that

$$\sum_i \mu_0\left(I_i\right) = \sum_{i,j} \mu_0\left(I_i \cap J_j\right) = \sum_j \mu_0\left(J_j\right). \tag{7.79}$$

Thus μ_0 is well defined, and it is finitely additive by construction.

It remains to show that if $\{I_j\}_1^\infty$ is a sequence of disjoint h-intervals with $\bigcup_1^\infty I_j \in \mathcal{A}$, then

$$\mu_0\left(\bigcup_1^\infty I_j\right) = \sum_1^\infty \mu_0\left(I_j\right). \tag{7.80}$$

Since $\bigcup_1^\infty I_j$ is a finite union of h-intervals, the sequence $\{I_j\}_1^\infty$ can be partitioned into finitely many subsequences such that the union of the intervals in each subsequence is a single h-interval. By considering each subsequence separately and using the finite additivity of μ_0, we may assume that $\bigcup_1^\infty I_j$ is an h-interval $I = (a, b]$.

The inequality $\geq$. For every n we have

$$\mu_0\left(I\right) = \mu_0\left(\bigcup_1^n I_j\right) + \mu_0\left(I - \bigcup_1^n I_j\right) \geq \mu_0\left(\bigcup_1^n I_j\right) = \sum_1^n \mu_0\left(I_j\right). \tag{7.81}$$

Letting $n \to \infty$, we obtain $\mu_0\left(I\right) \geq \sum_1^\infty \mu_0\left(I_j\right)$.

The inequality $\leq$. To prove the reverse inequality, let us suppose first that a and b are finite, and let us fix $\varepsilon > 0$. Since F is right continuous, there exists $\delta > 0$ such that

$$F\left(a + \delta\right) - F\left(a\right) < \varepsilon, \tag{7.82}$$

and, if $I_j = (a_j, b_j]$, for each j there exists $\delta_j > 0$ such that

$$F\left(b_j + \delta_j\right) - F\left(b_j\right) < \varepsilon 2^{-j}. \tag{7.83}$$

The open intervals $(a_j, b_j + \delta_j)$ cover the compact set $[a + \delta, b]$, so there is a finite subcover. By discarding any $(a_j, b_j + \delta_j)$ that is contained in a larger one and relabeling the index j, we may assume that

- the intervals $\left(a_1, b_1 + \delta_1\right), \ldots, \left(a_N, b_N + \delta_N\right)$ cover $[a + \delta, b]$,
- $b_j + \delta_j \in \left(a_{j+1}, b_{j+1} + \delta_{j+1}\right)$ for $j = 1, \ldots, N - 1$.

But then

$$\begin{aligned}
\mu_0(I) &< F(b) - F(a+\delta) + \varepsilon \\
&\leq F(b_N + \delta_N) - F(a_1) + \varepsilon \\
&= F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} [F(a_{j+1}) - F(a_j)] + \varepsilon \\
&\leq F(b_N + \delta_N) - F(a_N) + \sum_1^{N-1} [F(b_j + \delta_j) - F(a_j)] + \varepsilon \\
&< \sum_1^{N} \left[F(b_j) + \varepsilon 2^{-j} - F(a_j)\right] + \varepsilon \\
&< \sum_1^{\infty} \mu_0(I_j) + 2\varepsilon. \qquad (7.84)
\end{aligned}$$

Since ε is arbitrary, we are done when a and b are finite. If $a = -\infty$, for any $M < \infty$ the intervals $(a_j, b_j + \delta_j)$ cover $[-M, b]$, so the same reasoning gives

$$F(b) - F(-M) \leq \sum_1^{\infty} \mu_0(I_j) + 2\varepsilon, \qquad (7.85)$$

whereas if $b = \infty$, for any $M < \infty$ we likewise obtain

$$F(M) - F(a) \leq \sum_1^{\infty} \mu_0(I_j) + 2\varepsilon. \qquad (7.86)$$

The desired result then follows by letting $\varepsilon \to 0$ and $M \to \infty$. ■

The next result is a pay-off of our hard work in the previous subsections.

Theorem 7.7

If $F : \mathbb{R} \to \mathbb{R}$ is a non-decreasing and right-continuous function, there is a unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a, b]) = F(b) - F(a)$ for all a, b. If G is another such function, then $\mu_F = \mu_G$ if and only if $F - G$ is constant.

Conversely, if μ is a Borel measure on $\mathbb{R}$ that is finite on all bounded Borel sets and we define

$$F(x) = \begin{cases} \mu((0, x]), & \text{if } x > 0; \\ 0, & \text{if } x = 0; \\ -\mu((x, 0]), & \text{if } x < 0; \end{cases} \qquad (7.87)$$

then F is non-decreasing, right-continuous and $\mu = \mu_F$.

Proof. Each such F induces a pre-measure on $\mathcal{A}$ (finite disjoint unions of h-intervals) by Lemma 7.6. It is clear that F and G induce the same pre-measure if and only if $F - G$ is

constant. These pre-measures are σ-finite since

$$\mathbb{R} = \bigcup_{j=1}^{\infty} (j, j+1]. \tag{7.88}$$

The first assertion then follows from Theorem 7.4 and Lemma 7.5.

For the second part one can verify that F defined by (7.87) is non-decreasing and right-continuous, and that $\mu = \mu_F$ on $\mathcal{A}$. Hence $\mu = \mu_F$ on $\mathcal{B}_{\mathbb{R}}$ by the uniqueness part of Theorem 7.4 ■

The theory developed in the previous subsections gives, for each non-decreasing and right-continuous $F : \mathbb{R} \to \mathbb{R}$, not only a Borel measure μ_F but a complete measure $\overline{\mu_F}$ on a σ-algebra that contains $\mathcal{B}_{\mathbb{R}}$. In fact, $\overline{\mu_F}$ is just the completion of μ (see Exercise 7.3). This complete measure is usually referred to as the *Lebesgue–Stieltjes measure* associated to F. The Lebesgue measure on $\mathbb{R}$ is simply μ_F in the particular case of $F(x) = x$.

§8. 21 September 2026

We now discuss three informal heuristics, due to Littlewood, that summarize the "philosophy" of Lebesgue measure and integration on $\mathbb{R}^d$. Each heuristic is deliberately vague (hence the scare quotes around "nearly"); the content of this lecture is to make each one precise and prove it.

Littlewood's three principles.

(a) Every measurable set is "nearly" a finite union of cubes.

(b) Every measurable function is "nearly" continuous.

(c) Every pointwise convergent sequence of measurable functions is "nearly" uniformly convergent.

Remark 8.1. Principle (a) is exactly Proposition 6.4(ii), proved in the previous lecture: for E measurable with $m(E) < \infty$ and any $\varepsilon > 0$, there is a finite almost-disjoint union of cubes F' with $m(E \triangle F') \leq \varepsilon$.

We now turn to Principles (c) and (b), in that order (as the proof of Principle (b) will make use of Principle (c)).

§8.i. Egorov's theorem

Theorem 8.1 (Egorov's theorem)

Let $E \subseteq \mathbb{R}^d$ be measurable with $m(E) < \infty$, and let $(f_k)_{k \geq 1}$ be a sequence of measurable functions on E, finite-valued a.e., such that $f_k \to f$ pointwise a.e. on E, for some measurable f finite a.e. on E. Then for every $\varepsilon > 0$ there is a closed set $F_\varepsilon \subseteq E$ such that

$$m(E - F_\varepsilon) < \varepsilon \qquad \text{and} \qquad f_k \to f \text{ uniformly on } F_\varepsilon. \tag{8.1}$$

Proof. By discarding a null set from E (which affects no measure computation below), we may assume WLOG that $f_k(x) \to f(x)$ for *every* $x \in E$. For $n, k \geq 1$, define

$$E_{n,k} := \left\{ x \in E : |f_j(x) - f(x)| \leq \frac{1}{n} \text{ for all } j \geq k \right\}. \tag{8.2}$$

For fixed n,

- $E_{n,k}$ is increasing in k (a larger k is a weaker requirement, so $E_{n,k} \subseteq E_{n,k+1}$),
- measurable (an intersection of preimages under measurable functions),
- and $\bigcup_{k=1}^{\infty} E_{n,k} = E$: since $f_j(x) \to f(x)$ for every $x \in E$, so for any n, the condition $|f_j(x) - f(x)| \leq \frac{1}{n}$ holds for all sufficiently large j.

Now, continuity of measure from below gives $m(E_{n,k}) \to m(E)$ as $k \to \infty$,

$$\lim_{k\to\infty} m\left(E_{n,k}\right) = m\left(E\right) \implies \lim_{k\to\infty} m\left(E\right) - m\left(E_{n,k}\right) = 0 \implies \lim_{k\to\infty} m\left(E - E_{n,k}\right), \tag{8.3}$$

where the last implication follows since $m\left(E\right) < \infty$ and Corollary 6.3. Choose k_n such that

$$m(E - E_{n,k_n}) < \frac{\varepsilon}{2^{n+1}}. \tag{8.4}$$

Define

$$A_\varepsilon := \bigcap_{n=1}^{\infty} E_{n,k_n}. \tag{8.5}$$

Then $E - A_\varepsilon = \bigcup_{n=1}^{\infty}(E - E_{n,k_n})$, so by countable subadditivity,

$$m(E - A_\varepsilon) \;\le\; \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \frac{\varepsilon}{2}. \tag{8.6}$$

Now, we'll show that $f_j \to f$ uniformly on A_ε. Given $\delta > 0$, choose n with $\frac{1}{n} < \delta$. For every $x \in A_\varepsilon$, $x \in E_{n,k_n}$, so

$$|f_j(x) - f(x)| \le \frac{1}{n} < \delta \quad \text{for } j \ge k_n. \tag{8.7}$$

Since k_n does not depend on x, this is uniform convergence on A_ε.

Now, by Lemma 6.1, choose a closed set $F_\varepsilon \subseteq A_\varepsilon$ with $m(A_\varepsilon - F_\varepsilon) \le \varepsilon/2$. Then

$$m(E - F_\varepsilon) \;\le\; m(E - A_\varepsilon) + m(A_\varepsilon - F_\varepsilon) \;\le\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \tag{8.8}$$

and $f_k \to f$ uniformly on $F_\varepsilon \subseteq A_\varepsilon$, since uniform convergence on a set persists on every subset. ■

§8.ii. Lusin's theorem

Definition 8.1 (Step function)**.** A **step function** on $\mathbb{R}^d$ is a finite linear combination of characteristic functions of cubes:

$$\psi = \sum_{j=1}^{N} a_j\, \chi_{Q_j}, \tag{8.9}$$

where $Q_1, \ldots, Q_N$ are cubes (not necessarily disjoint) and $a_1, \ldots, a_N \in \mathbb{R}$.

Lemma 8.2 (Approximation of characteristic functions by step functions)

Let $E \subseteq \mathbb{R}^d$ be measurable and $m\left(E\right) < \infty$. For every $\varepsilon > 0$, there is a step function ψ such that

$$m\left(\left\{x \in \mathbb{R}^d : \psi(x) \ne \chi_E(x)\right\}\right) \le \varepsilon. \tag{8.10}$$

Proof. By Proposition 6.4(ii), choose a finite almost disjoint union of cubes

$$F' = \bigcup_{j=1}^{N} Q_j, \qquad \text{with } m(E \triangle F') \leq \varepsilon. \tag{8.11}$$

Let $\psi := \sum_{j=1}^{N} \chi_{Q_j}$, a step function. For $x \in \bigcup_{j=1}^{N} Q_j^\circ$, since it's a disjoint union

$$\psi(x) = 1 = \chi_{F'}(x). \tag{8.12}$$

For $x \notin F$, $x \notin Q_j$ for any j, so

$$\psi(x) = 0 = \chi_{F'}(x). \tag{8.13}$$

Therefore, $\psi = \chi_{F'}$ on B^c, where $B = \bigcup_{j=1}^{N} (\partial Q_j)$ is the union of the boundaries of Q_j. Hence,

$$\left\{x \in \mathbb{R}^d : \chi_E(x) \neq \psi(x)\right\} \cap B^c = \left\{x \in \mathbb{R}^d : \chi_E(x) \neq \chi_{F'}(x)\right\} \cap B^c = \left(E \triangle F'\right) \cap B^c. \tag{8.14}$$

Now, using the fact that $m(X) = m(X \cap B) + m(X \cap B^c)$ (this is a consequence of additivity) for measurable sets X and B, and if B is a null-set

$$m(X) = m(X \cap B) + m(X \cap B^c) = m(X \cap B^c). \tag{8.15}$$

Therefore,

$$\begin{aligned} m\left(\left\{x \in \mathbb{R}^d : \chi_E(x) \neq \psi(x)\right\}\right) &= m\left(\left\{x \in \mathbb{R}^d : \chi_E(x) \neq \psi(x)\right\} \cap B^c\right) \\ &= m\left(\left(E \triangle F'\right) \cap B^c\right) \\ &= m\left(E \triangle F'\right) \leq \varepsilon. \end{aligned} \tag{8.16}$$

■

Proposition 8.3 (Approximation of measurable functions by step functions, a.e.)

Let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable and finite-valued a.e. Then there is a sequence of step functions $(\psi_k)_{k \geq 1}$ such that $\psi_k \to f$ a.e. on $\mathbb{R}^d$.

Proof. By the approximation of measurable functions by simple functions (Theorem 2.3 applied to f^+ and f^- separately), there is a sequence of simple functions $\phi_n \to f$ everywhere. There is no loss of generality if we replace ϕ_n with $\varphi_n := \phi_n \cdot \chi_{B_n}$ (where $B_n = \overline{B(\mathbf{0}, n)}$ is the closed ball of radius n centered at the origin). This does not affect pointwise convergence, as $\chi_{B_n}(x) = 1$ for $n \geq |x|$.

Therefore, there is a sequence of simple functions $\varphi_j \to f$ pointwise, each of the form

$$\varphi_j = \sum_{i=1}^{n_j} a_{ij} \chi_{E_{ij}} \tag{8.17}$$

with every E_{ij} bounded and measurable. By Proposition 8.2, for each i, j choose a step function η_{ij} with

$$m\left(\{\eta_{ij} \neq \chi_{E_{ij}}\}\right) \leq \frac{1}{2^j \; n_j(1+|a_{ij}|)}, \tag{8.18}$$

and set

$$\psi_j = \sum_{i=1}^{n_j} a_{ij}\, \eta_{ij} \tag{8.19}$$

a step function. Let

$$Y_j := \bigcup_{i=1}^{n_j} \{\eta_{ij} \neq \chi_{E_{ij}}\}. \tag{8.20}$$

Then we have

$$m(Y_j) \leq \sum_{i=1}^{n_j} m\left(\{\eta_{ij} \neq \chi_{E_{ij}}\}\right) \leq \sum_{i=1}^{n_j} \frac{1}{2^j \; n_j(1+|a_{ij}|)} \leq \frac{1}{2^j}. \tag{8.21}$$

Outside Y_j, $\eta_{ij} = \chi_{E_{ij}}$ for every i, so $\psi_j(x) = \varphi_j(x)$ for $x \notin Y_j$. Since $\sum_{j=1}^{\infty} m(Y_j) < \infty$, the Borel–Cantelli lemma gives

$$m\left(Y^*\right) = m\left(\{x : x \in Y_j \text{ for infinitely many } j\}\right) = 0. \tag{8.22}$$

If x lies in only finitely many Y_j, so $\psi_j(x) = \varphi_j(x)$ for all sufficiently large j. Since $\varphi_j(x) \to f(x)$, we can conclude that $\psi_j(x) \to f(x)$ holds for every $x \notin Y^*$, i.e. almost everywhere, since $m(Y^*) = 0$. ∎

Theorem 8.4 (Lusin's theorem)

Let $E \subseteq \mathbb{R}^d$ be measurable with $m(E) < \infty$, and let $f : E \to \mathbb{R}$ be measurable and finite-valued. Then for every $\varepsilon > 0$ there is a closed set $F_\varepsilon \subseteq E$ such that

$$m(E - F_\varepsilon) < \varepsilon \qquad \text{and} \qquad f|_{F_\varepsilon} \text{ is continuous.} \tag{8.23}$$

Remark 8.2. The conclusion concerns the *restriction* $f|_{F_\varepsilon}$, viewed as a function $F_\varepsilon \to \mathbb{R}$ in its own right (with the subspace topology); it does *not* assert that f, as a function on E, is continuous *at* the points of F_ε, which is a much stronger and generally false statement. For example, $f = \chi_{\mathbb{Q}\cap[0,1]}$ on $[0,1]$ is discontinuous at *every* point of $[0,1]$, yet f restricted to the set of irrationals (of full measure) is identically 0, hence (trivially) continuous.

Proof. Extend f to all of $\mathbb{R}^d$ by setting $f \equiv 0$ outside E; this does not affect measurability. By Proposition 8.3, choose step functions $\psi_k \to f$ a.e. on $\mathbb{R}^d$. Each step function

$$\psi_k = \sum_{i=1}^{n_i} a_{ik} \chi_{Q_{ik}} \tag{8.24}$$

is continuous at every point outside of the boundary

$$B_k = \bigcup_i^{n_i} \partial Q_{ik}, \tag{8.25}$$

which a closed set with $m(B_k) = 0$.

By Egorov's theorem (applied on E, using $m(E) < \infty$), choose a closed set $\widetilde{F} \subseteq E$ with $m(E - \widetilde{F}) \leq \varepsilon/2$ such that $\psi_k \to f$ uniformly on $\widetilde{F}$. Let

$$G := \widetilde{F} - \bigcup_{k=1}^{\infty} B_k. \tag{8.26}$$

Since

$$m\left(\bigcup_{k=1}^{\infty} B_k\right) \leq \sum_{k=1}^{\infty} m(B_k) = 0, \tag{8.27}$$

we have $m(\widetilde{F} - G) = 0$, so

$$m(E - G) \leq m(E - \widetilde{F}) + m(\widetilde{F} - G) \leq \frac{\varepsilon}{2}. \tag{8.28}$$

On G, every ψ_k is continuous (having removed all cube boundaries), and $\psi_k \to f$ uniformly on $G \subseteq \widetilde{F}$; since a uniform limit of continuous functions is continuous, $f|_G$ is continuous.

Finally, G need not be closed. By Lemma 6.1, choose a closed set $F_\varepsilon \subseteq G$ with $m(G - F_\varepsilon) \leq \frac{\varepsilon}{2}$; then

$$m(E - F_\varepsilon) \leq m(E - G) + m(G - F_\varepsilon) \leq \varepsilon, \tag{8.29}$$

and $f|_{F_\varepsilon}$, being the restriction of the continuous function $f|_G$ to a subset, is continuous. ■

§8.iii. Exercises

Exercise 8.1. Let $f : [0,1] \to [0,1]$ be measurable, one-to-one, and onto, satisfying the following condition:

$$\forall N \subseteq [0,1]\,,\ m\,(N) = 0 \implies m\,(f\,(N)) = 0. \tag{8.30}$$

Prove that $f^{-1} : [0,1] \to [0,1]$ is measurable.

Solution. By Lusin's theorem, there is a closed set $F_n \subseteq [0,1]$ such that

$$m\,([0,1] - F_n) \leq \frac{1}{n}, \quad \text{and } f|_{F_n} \text{ is continuous.} \tag{8.31}$$

Since each F_n is closed and bounded, they are all compact. Consider

$$N := [0,1] - \left(\bigcup_{n=1}^{\infty} F_n\right). \tag{8.32}$$

Then $N \subseteq [0,1] - F_n$ for each n. Hence,

$$m(N) \leq m([0,1] - F_n) \leq \frac{1}{n}, \tag{8.33}$$

for each n, implying that $m(N) = 0$. By definition of measurable functions, f^{-1} is measurable if and only if

$$\left(f^{-1}\right)^{-1}((a,\infty)) = f((a,\infty) \cap [0,1]) \tag{8.34}$$

is Lebesgue measurable for every $a \in \mathbb{R}$. We write

$$(a,\infty) \cap [0,1] = (a,\infty) \cap \left(\bigcup_{n=1}^{\infty} F_n \cup N\right) = \bigcup_{n=1}^{\infty} ((a,\infty) \cap F_n) \cup ((a,\infty) \cap N) \tag{8.35}$$

Therefore,

$$f((a,\infty) \cap [0,1]) = \bigcup_{n=1}^{\infty} f((a,\infty) \cap F_n) \cup f((a,\infty) \cap N) \tag{8.36}$$

$(a,\infty) \cap N \subseteq N$, so it's a nullset. Therefore, $f((a,\infty) \cap N)$ has measure 0, by hypothesis. So it's Lebesgue measurable. For other's, we'll use the following result from topology:

> If X is a compact space and Y is a Hausdorff space, and $g : X \to Y$ is a bijective continuous map, then g is a homeomorphism.

Now, $f|_{F_n}$: $F_n \to f(F_n)$ is a continuous bijective map from a compact space to a Hausdorff space, so it's a homeomorphism. Therefore, it maps (relative) open sets to (relative) open sets. This proves that $f((a,\infty) \cap F_n)$ is open in $f(F_n)$, i.e. it's an open set in $\mathbb{R}$ intersected with $f(F_n)$ (which is compact). So it's also Lebesgue measurable.

Since all the parts in the RHS of (8.36) are Lebesgue measurable, we can conclude that f^{-1} is measurable. ■

§9. 23 September 2026

Having developed general measure theory, the construction of Lebesgue measure, and the basic convergence theorems (culminating in Littlewood's three principles), we now turn to the theory of *function spaces*. The first, and most fundamental, family of examples is the L^p spaces.

§9.i. L^p spaces

Throughout, $(\Omega, \mathcal{M}, \mu)$ denotes a measure space: a set Ω, a σ-algebra $\mathcal{M}$ of subsets of Ω, and a measure μ on $\mathcal{M}$. We consider measurable functions $f : \Omega \to \mathbb{R}$ (or $\Omega \to \mathbb{C}$).

Definition 9.1 (L^p norm)**.** For a measurable function f and $1 \le p < \infty$, define

$$\|f\|_p = \left(\int_\Omega |f(x)|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} \in [0, \infty]. \tag{9.1}$$

For $p = \infty$, define the *essential supremum* of f:

$$\|f\|_\infty = \inf \{ \alpha \ge 0 \ : \ |f(x)| \le \alpha \text{ for } \mu\text{-a.e. } x \} =: \operatorname{ess\,sup} f. \tag{9.2}$$

Remark 9.1 (Why "essential")**.** The essential supremum is designed to be insensitive to changes of f on a set of measure zero, unlike the ordinary supremum. If f equals some bounded function except on a null set, where it is altered arbitrarily (even made unbounded there), $\|f\|_\infty$ is unaffected: it is the smallest α such that $|f| \le \alpha$ holds *almost everywhere*, not everywhere.

Remark 9.2 (Equivalent formulation)**.** One can equivalently write

$$\|f\|_\infty = \inf_{\substack{N \subseteq \Omega \\ \mu(N) = 0}} \sup_{x \in \Omega - N} |f(x)| \,, \tag{9.3}$$

i.e. the infimum, over all null sets N, of the ordinary supremum of $|f|$ on the complement of N.

Remark 9.3. The cases $0 < p < 1$ can also be defined formally by the same integral formula, but $\|\cdot\|_p$ fails to be a norm for such p (the triangle inequality fails), so these are not normed spaces in the usual sense; we will not use them.

§9.ii. Basic Properties of the L^p spaces

Proposition 9.1 (Homogeneity)

For any scalar c and $1 \le p \le \infty$,

$$\|cf\|_p = |c| \, \|f\|_p \,. \tag{9.4}$$

Proof. For $1 \le p < \infty$, pull $|c|^p$ out of the integral and take the p-th root:

$$\|cf\|_p = \left(\int_\Omega |c|^p \, |f|^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} = |c| \left(\int_\Omega |f|^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} = |c| \, \|f\|_p \,. \tag{9.5}$$

For $p = \infty$, $|cf(x)| \le \alpha$ a.e. if and only if $|f(x)| \le \alpha/|c|$ a.e. (for $c \neq 0$), so the essential suprema scale accordingly. The case $c = 0$ is immediate. ∎

Proposition 9.2 (Finiteness is preserved under addition)

If $\|f\|_p < \infty$ and $\|g\|_p < \infty$ for some $1 \le p \le \infty$, then $\|f + g\|_p < \infty$.

Proof. **Case 1:** $p = \infty$**.** If $|f(x)| \le \alpha$ a.e. and $|g(x)| \le \beta$ a.e., then $|f(x) + g(x)| \le \alpha + \beta$ a.e., so $\|f + g\|_\infty \le \alpha + \beta < \infty$.

Case $1 \le p < \infty$**.** Pointwise,

$$|f(x) + g(x)|^p \;\le\; (2 \max\{|f(x)|, |g(x)|\})^p = 2^p \max\{|f(x)|, |g(x)|\}^p \;\le\; 2^p (\, |f(x)|^p + |g(x)|^p \,). \tag{9.6}$$

Integrating,

$$\int_\Omega |f + g|^p \, \mathrm{d}\mu \;\le\; 2^p \left(\int_\Omega |f|^p \, \mathrm{d}\mu + \int_\Omega |g|^p \, \mathrm{d}\mu\right) < \infty. \tag{9.7}$$

∎

Remark 9.4. This is only a finiteness statement, not yet the (much sharper) triangle inequality $\|f + g\|_p \le \|f\|_p + \|g\|_p$, which will be proved later.

Proposition 9.3 (Vanishing)

For $1 \le p < \infty$, $\|f\|_p = 0$ if and only if $f = 0$ μ-almost everywhere.

Proof. This follows immediately from Proposition 2.9. ∎

Remark 9.5. Homogeneity and finiteness-under-addition show that for fixed $1 \le p \le \infty$, the collection $\{f : \|f\|_p < \infty\}$ is closed under scalar multiplication and addition, hence is a vector space. Once the triangle inequality is established, $\|\cdot\|_p$ will be a genuine norm on the space of *equivalence classes* of functions equal a.e., since as it stands $\|f\|_p = 0$ only forces $f = 0$ a.e., not $f \equiv 0$ (i.e. it is a priori only a *seminorm* before this identification).

§9.iii. Young's Inequality

Definition 9.2 (Conjugate exponents)**.** Two numbers $1 \le p, q \le \infty$ are *conjugate exponents* if $\frac{1}{p} + \frac{1}{q} = 1$ (with the convention $\frac{1}{\infty} = 0$, so $p = 1$ is conjugate to $q = \infty$).

Lemma 9.4 (Young's inequality)

Let $1 < p, q < \infty$ be conjugate exponents. For all real numbers $a, b \geq 0$,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \tag{9.8}$$

with equality if and only if $a^p = b^q$.

Proof. If $b = 0$ the inequality is trivial, so assume $b > 0$. Define, for $x \geq 0$,

$$\varphi(x) = \frac{x^p}{p} + \frac{b^q}{q} - xb. \tag{9.9}$$

Then $\varphi'(x) = x^{p-1} - b$, which is negative for $x < b^{\frac{1}{p-1}}$, zero at $x_0 = b^{\frac{1}{p-1}}$, and positive for $x > b^{\frac{1}{p-1}}$; so φ attains its global minimum on $[0, \infty)$ at x_0. Compute, using $q = \frac{p}{p-1}$ (equivalently $(p-1)q = p$),

$$\varphi(x_0) = \frac{x_0^p}{p} + \frac{b^q}{q} - x_0 b = \frac{b^{\frac{p}{p-1}}}{p} + \frac{b^q}{q} - b^{\frac{1}{p-1}} b \tag{9.10}$$

$$= \frac{b^q}{p} + \frac{b^q}{q} - b^q \tag{9.11}$$

$$= b^q \left(\frac{1}{p} + \frac{1}{q} - 1 \right) = 0. \tag{9.12}$$

Hence $\varphi(x) \geq \varphi(x_0) = 0$ for all $x \geq 0$, i.e. $xb \leq \frac{x^p}{p} + \frac{b^q}{q}$ for all $x \geq 0$. Taking $x = a$ gives the claimed inequality. Equality holds exactly when $a = x_0 = b^{\frac{1}{p-1}}$,

$$a^{p-1} = b \iff a^p = ab = b^{\frac{1}{p-1}+1} = b^q. \tag{9.13}$$

■

§9.iv. Hölder's Inequality

Theorem 9.5 (Hölder's inequality)

Let $1 \leq p, q, r \leq \infty$ satisfy

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r}. \tag{9.14}$$

If f, g are measurable functions,

$$\|fg\|_r \leq \|f\|_p \, \|g\|_q . \tag{9.15}$$

When the RHS of (9.15) is finite, and $1 < p, q < \infty$, equality holds if and only if $f = 0$ a.e., or $g = 0$ a.e., or there is $\alpha > 0$ with $|g(x)|^q = \alpha \, |f(x)|^p$ for a.e. x.

Remark 9.6. The classical statement usually seen in textbooks is the case $r = 1$, where p and

q are conjugate exponents. The general statement above (with an arbitrary target exponent r) reduces to that case, as shown below.

Remark 9.7 (Why this relation among exponents)**.** The relation $\frac{1}{p}+\frac{1}{q}=\frac{1}{r}$ is exactly what makes (9.15) consistent with respect to dilation of the space. Suppose we put no restriction on the exponents p, q, r yet. We want to have a universal inequality

$$\|fg\|_r \;\leq\; C\,\|f\|_p\,\|g\|_q\,, \tag{9.16}$$

where the constant C does not depend on f, g. Suppose we dilated the space $\Omega=\mathbb{R}^d$ by a positive factor $\lambda>0$; in other words, we replace f, g with

$$f_\lambda(x)=f(\lambda x)\,,\quad g_\lambda(x)=g(\lambda x)\,. \tag{9.17}$$

Then (9.16) should hold for f_λ, g_λ as well, i.e.

$$\|f_\lambda g_\lambda\|_r \;\leq\; C\,\|f_\lambda\|_p\,\|g_\lambda\|_q\,. \tag{9.18}$$

Now,

$$\|f_\lambda\|_p=\left(\int_{\mathbb{R}^d}|f(\lambda x)|^p\,\mathrm{d}x\right)^{\frac{1}{p}}=\left(\int_{\mathbb{R}^d}|f(u)|^p\,\lambda^{-d}\,\mathrm{d}u\right)^{\frac{1}{p}}=\lambda^{-\frac{d}{p}}\,\|f\|_p\,, \tag{9.19}$$

where we have invoked the change of variables $u=\lambda x$, so that $\mathrm{d}x=\lambda^{-d}\,\mathrm{d}u$. In a similar manner as (9.19), we get

$$\|g_\lambda\|_q=\lambda^{-\frac{d}{q}}\,\|g\|_q\,,\quad \|f_\lambda g_\lambda\|_r=\lambda^{-\frac{d}{r}}\,\|fg\|_r\,. \tag{9.20}$$

Plugging everything back to (9.18), we have

$$\lambda^{-\frac{d}{r}}\,\|fg\|_r\leq C\cdot\lambda^{-\frac{d}{p}}\,\|f\|_p\cdot\lambda^{-\frac{d}{q}}\,\|g\|_q\,. \tag{9.21}$$

After rearranging,

$$\|fg\|_r\leq C\cdot\lambda^{d\left(\frac{1}{r}-\frac{1}{p}-\frac{1}{q}\right)}\,\|f\|_p\,\|g\|_q\,, \tag{9.22}$$

which should hold for any $\lambda>0$.

- If $\frac{1}{r}-\frac{1}{p}-\frac{1}{q}>0$, then as $\lambda\to 0$, the RHS approaches 0. This implies $\|fg\|_r\leq 0$, which is obviously false for general measurable functions f, g.
- If $\frac{1}{r}-\frac{1}{p}-\frac{1}{q}<0$, then as $\lambda\to\infty$, the RHS approaches 0. This also implies $\|fg\|_r\leq 0$.

Therefore, the only way an inequality like $\|fg\|_r \;\leq\; C\,\|f\|_p\,\|g\|_q$ can make sense if $\frac{1}{r}-\frac{1}{p}-\frac{1}{q}=0$. In this case, the constant C is actually 1, as we'll prove now.

Proof of Hölder's inequality. If $\|f\|_p=0$ or $\|g\|_q=0$, then (by the Proposition 9.3) $f=0$ or $g=0$ a.e., so $fg=0$ a.e. and both sides of (9.15) are 0. So assume $\|f\|_p, \|g\|_q>0$.

If $\|f\|_p=\infty$ or $\|g\|_q=\infty$, (9.15) is trivially true (right side infinite). So assume also $0<\|f\|_p, \|g\|_q<\infty$.

Case 1: $r = \infty$ (hence $p = q = \infty$). If $|f| \le \alpha$ a.e. and $|g| \le \beta$ a.e., then $|fg| \le \alpha\beta$ a.e., so $\|fg\|_\infty \le \|f\|_\infty \|g\|_\infty$.

Case 2: $p = \infty$ or $q = \infty$ with $r < \infty$, say $q = \infty$ (so $r = p$). Then $|g(x)| \le \|g\|_\infty$ a.e., so

$$\int_\Omega |fg|^p \, \mathrm{d}\mu \le \int_\Omega |f|^p \, \|g\|_\infty \, \mathrm{d}\mu = \|g\|_\infty^p \int_\Omega |f|^p \, \mathrm{d}\mu, \tag{9.23}$$

and taking p-th roots gives $\|fg\|_p \le \|f\|_p \|g\|_\infty$.

Case 3: $r = 1$, $1 < p, q < \infty$ conjugate. It remains to prove

$$\int_\Omega |fg| \, \mathrm{d}\mu \le \|f\|_p \|g\|_q . \tag{9.24}$$

If $\|f\|_p = 0$ or $\|g\|_q = 0$ this is trivial as noted above, so assume both are finite and positive. Apply Young's inequality pointwise to $a = \frac{|f(x)|}{\|f\|_p}$ and $b = \frac{|g(x)|}{\|g\|_q}$:

$$\frac{|f(x)|}{\|f\|_p} \cdot \frac{|g(x)|}{\|g\|_q} \;\le\; \frac{1}{p}\left(\frac{|f(x)|}{\|f\|_p}\right)^p + \frac{1}{q}\left(\frac{|g(x)|}{\|g\|_q}\right)^q . \tag{9.25}$$

Integrating both sides over Ω,

$$\frac{1}{\|f\|_p \|g\|_q} \int_\Omega |fg| \, \mathrm{d}\mu \;\le\; \frac{1}{p} \cdot \frac{1}{\|f\|_p^p} \int_\Omega |f|^p \, \mathrm{d}\mu + \frac{1}{q} \cdot \frac{1}{\|g\|_q^q} \int |g|^q \, \mathrm{d}\mu = \frac{1}{p} + \frac{1}{q} = 1, \tag{9.26}$$

using the definitions of $\|f\|_p, \|g\|_q$. Multiplying through by $\|f\|_p \|g\|_q$ gives $\int_\Omega |fg| \, \mathrm{d}\mu \le \|f\|_p \|g\|_q$, as required.

Case 4: $r > 1$. Suppose $1 < r < \infty$ and $1 \le p, q < \infty$. Let

$$F = |f|^r , \quad G = |g|^r , \quad P = \frac{p}{r}, \quad Q = \frac{q}{r}. \tag{9.27}$$

Then

$$\frac{1}{P} + \frac{1}{Q} = \frac{r}{p} + \frac{r}{q} = r\left(\frac{1}{p} + \frac{1}{q}\right) = 1, \tag{9.28}$$

so P, Q are conjugate exponents. Applying the case $r = 1$ applied to F, G, P, Q,

$$\int_\Omega FG \, \mathrm{d}\mu \le \left(\int_\Omega F^P \, \mathrm{d}\mu\right)^{\frac{1}{P}} \left(\int_\Omega G^Q \, \mathrm{d}\mu\right)^{\frac{1}{Q}} , \tag{9.29}$$

i.e.

$$\int_\Omega |f|^r \, |g|^r \, \mathrm{d}\mu \le \left(\int_\Omega |f|^p \, \mathrm{d}\mu\right)^{\frac{r}{p}} \left(\int_\Omega |g|^q \, \mathrm{d}\mu\right)^{\frac{r}{q}} , \tag{9.30}$$

and raising both sides to the power $\frac{1}{r}$ gives exactly $\|fg\|_r \le \|f\|_p \|g\|_q$.

Equality. Equality in the integrated Young's inequality forces pointwise equality a.e. (since the

difference of the two sides is a nonnegative function integrating to 0), i.e.

$$\left(\frac{|f(x)|}{\|f\|_p}\right)^p = \left(\frac{|g(x)|}{\|g\|_q}\right)^q \qquad \text{a.e.,} \tag{9.31}$$

which is exactly $|g(x)|^q = \alpha\ |f(x)|^p$ a.e. for $\alpha = \|g\|_q^q / \|f\|_p^p > 0$. ■

Remark 9.8. $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ is only a *necessary* condition, extracted purely from the dilation symmetry of Lebesgue measure on $\mathbb{R}^d$ (that $m(\lambda E) = \lambda^d m(E)$); it says nothing about whether the inequality actually holds once the relation is satisfied. Here it does hold (Hölder's inequality), but the scaling argument alone only rules out the impossible relations, narrowing down the search before any real analytic work is done. The moral: any inequality intended to hold globally on $\mathbb{R}^d$ must respect the symmetries (here, dilation invariance) of the underlying measure.

Remark 9.9 (Contrast: finite measure spaces)**.** On a measure space with $\mu(\Omega) = 1$ (a probability space), there is no dilation symmetry to exploit, and indeed no relation among exponents is forced in the same way. In fact, on such a space the L^p norms are *nested*: if $1 \le a < b \le \infty$, then

$$\|f\|_a \ \le\ \|f\|_b \qquad \text{for every measurable } f. \tag{9.32}$$

Assume $b < \infty$ (the case $b = \infty$ is similar). Apply Hölder's inequality to $|f|^a \cdot 1$ with exponents $c = b/a$ (so $c > 1$) and its conjugate c', where $1/c + 1/c' = 1$:

$$\int_\Omega |f|^a \cdot 1 \, \mathrm{d}\mu \ \le\ \left(\int_\Omega |f|^{ac} \, \mathrm{d}\mu\right)^{1/c} \left(\int_\Omega 1^{c'} \, \mathrm{d}\mu\right)^{1/c'} = \|f\|_b^a \cdot \mu(\Omega)^{1/c'}. \tag{9.33}$$

Since $\mu(\Omega) = 1$, the second factor is 1, so $\|f\|_a^a = \int |f|^a \, \mathrm{d}\mu \le \|f\|_b^a$, i.e. $\|f\|_a \le \|f\|_b$.

Therefore, if $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, and $q' > q$, we have

$$\|fg\|_r \le \|f\|_p \, \|g\|_q \le \|f\|_p \, \|g\|_{q'}\,, \tag{9.34}$$

so $\|fg\|_r \le \|f\|_p \, \|g\|_{q'}$, despite $\frac{1}{p} + \frac{1}{q'} \ne \frac{1}{r}$. Therefore, if we restrict our attention to a specific measure space, $\|fg\|_r \le \|f\|_p \, \|g\|_q$ may happen **without** $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. But for $\mathbb{R}^d$, the condition $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ is necessary.

Remark 9.10 (Another instance of the scaling heuristic)**.** As a second illustration of the scaling method, consider inequalities of the form

$$\|f\|_q \ \le\ C\ \|\nabla f\|_p\,, \qquad f \in C_c^\infty(\mathbb{R}^d) \tag{9.35}$$

(smooth, compactly supported functions), where ∇f is the gradient. The heuristic is: to make f large while keeping its support fixed, the function must vary quickly, forcing ∇f to be large; so a bound on the size of the gradient should control the size of the function itself, *provided* f is compactly supported (so it cannot simply be a large constant, which has zero gradient but unbounded L^q norm if the domain were unbounded).

If such an inequality holds for all $f \in C_c^\infty(\mathbb{R}^d)$ with C independent of f, then we need to have some condition on p, q, d. We argue similarly as before. For $\lambda > 0$ let

$$f_\lambda(x) = f(\lambda x). \tag{9.36}$$

Then

$$\nabla f_\lambda(x) = \lambda\,(\nabla f)(\lambda x) \tag{9.37}$$

As in the previous scaling computation,

$$\|f_\lambda\|_q = \lambda^{-\frac{d}{q}}\,\|f\|_q\,, \qquad \|\nabla f_\lambda\|_p = \lambda \cdot \lambda^{-\frac{d}{p}}\,\|\nabla f\|_p = \lambda^{1-\frac{d}{p}}\,\|\nabla f\|_p\,. \tag{9.38}$$

Applying the hypothesized inequality (9.35) to f_λ:

$$\lambda^{-\frac{d}{q}}\,\|f\|_q \;\le\; C\,\lambda^{1-\frac{d}{p}}\,\|\nabla f\|_p \qquad \text{for all } \lambda > 0. \tag{9.39}$$

After rearranging,

$$\|f\|_q \;\le\; C\,\lambda^{d\left(\frac{1}{q}+\frac{1}{d}-\frac{1}{p}\right)}\,\|\nabla f\|_p \qquad \text{for all } \lambda > 0. \tag{9.40}$$

As before, the only way this can hold for every $\lambda > 0$ is if the exponent of λ is 0, i.e.

$$\frac{1}{d} = \frac{1}{p} - \frac{1}{q}. \tag{9.41}$$

The exponent $q = \frac{dp}{d-p}$ determined this way is called the *Sobolev conjugate* of p (defined when $p < d$). That the Sobolev inequality **actually** holds with this exponent is a real theorem (the Sobolev embedding theorem), well beyond what the scaling argument alone establishes. The scaling argument only identifies the *unique candidate* exponent for which such an inequality could possibly be true, exactly as in the Hölder case above.

Remark 9.11 (General moral)**.** Every "global" inequality on $\mathbb{R}^d$, one asserted to hold for all functions in certain norms, with a universal constant, must be consistent with the dilation (and, more generally, translation and rotation) symmetries of Lebesgue measure. Checking this consistency via a one-parameter scaling substitution is a quick and standard first step, before attempting to prove the inequality itself, in identifying which exponent relations are even conceivable.

§9.v. Minkowski's Inequality

Theorem 9.6 (Minkowski's inequality)

Let $1 \le p \le \infty$ and f, g be measurable functions. Then

$$\|f+g\|_p \;\le\; \|f\|_p + \|g\|_p\,. \tag{9.42}$$

For $1 < p < \infty$, equality holds if and only if $g = \alpha f$ a.e. for some $\alpha \ge 0$, or $f = 0$ a.e., or

$g = 0$ a.e. For $p = 1$, equality holds if and only if $f(x)\overline{g(x)} \geq 0$ a.e. (in the real case: f and g have the same sign a.e., allowing either to vanish).

Proof. Finiteness of $\|f+g\|_p$ was already shown in Proposition 9.2.

Case 1: $p = 1$**.** Immediate from the pointwise triangle inequality $|f(x)+g(x)| \leq |f(x)|+|g(x)|$, integrated. Equality throughout requires pointwise equality a.e., which (for real or complex numbers) holds iff $f(x)$ and $g(x)$ are nonnegative multiples of one another at a.e. x (they point in the same direction in $\mathbb{C}$, equivalently have the same sign when real).

Case 2: $p = \infty$**.** If $|f| \leq \alpha$ a.e. and $|g| \leq \beta$ a.e., then $|f+g| \leq \alpha+\beta$ a.e., so $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

Case 3: $1 < p < \infty$**.** If $\|f+g\|_p = 0$ the inequality is trivial, so assume $\|f+g\|_p > 0$; we may also assume $\|f+g\|_p < \infty$ (already shown). Pointwise, using the triangle inequality,

$$|f(x)+g(x)|^p = |f(x)+g(x)| \cdot |f(x)+g(x)|^{p-1} \;\leq\; \big(\,|f(x)|+|g(x)|\,\big)\,|f(x)+g(x)|^{p-1}\,. \tag{9.43}$$

Integrating,

$$\int_\Omega |f+g|^p \,\mathrm{d}\mu \;\leq\; \int_\Omega |f|\;|f+g|^{p-1}\,\mathrm{d}\mu \;+\; \int_\Omega |g|\;|f+g|^{p-1}\,\mathrm{d}\mu. \tag{9.44}$$

Let $q = \frac{p}{p-1}$, the conjugate exponent to p. Apply Hölder's inequality (with exponents p, q) to each term on the right of (9.44), treating $|f+g|^{p-1}$ as the second factor:

$$\int_\Omega |f|\;|f+g|^{p-1}\,\mathrm{d}\mu \leq \|f\|_p \left(\int_\Omega |f+g|^{(p-1)q}\,\mathrm{d}\mu\right)^{\frac{1}{q}} \tag{9.45}$$

$$= \|f\|_p \left(\int_\Omega |f+g|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}\cdot\frac{p}{q}} \tag{9.46}$$

$$= \|f\|_p \;\|f+g\|_p^{p/q} \tag{9.47}$$

Similarly,

$$\int_\Omega |g|\;|f+g|^{p-1}\,\mathrm{d}\mu \leq \|g\|_p \;\|f+g\|_p^{p/q}\,. \tag{9.48}$$

Substituting into (9.44),

$$\|f+g\|_p^p \;\leq\; \big(\,\|f\|_p + \|g\|_p\,\big)\,\|f+g\|_p^{p/q}\,. \tag{9.49}$$

Since $0 < \|f+g\|_p < \infty$, divide both sides by $\|f+g\|_p^{p/q}$, and use

$$p - \frac{p}{q} = p\left(1-\frac{1}{q}\right) = p\cdot\frac{1}{p} = 1, \tag{9.50}$$

to arrive at

$$\|f+g\|_p \;=\; \|f+g\|_p^{p-p/q} \;\leq\; \|f\|_p + \|g\|_p\,. \tag{9.51}$$

Equality. Tracing the argument back, equality in Minkowski's inequality for $1 < p < \infty$ requires equality both in the pointwise triangle inequality $|f+g| \le |f| + |g|$ (a.e.) and in both applications of Hölder's inequality. The former forces f, g to have the same sign (or one to vanish) a.e. The latter, via the equality case of Hölder, forces $|f+g|^{(p-1)q} = |f+g|^p$ to be proportional to $|f|^p$ and to $|g|^p$ a.e., which (combined with the same-sign condition) reduces to $g = \alpha f$ a.e. for some $\alpha \ge 0$, or one of f, g vanishing a.e. ■

We have now shown that for $1 \le p \le \infty$, $\|\cdot\|_p$ satisfies:

- $\|cf\|_p = |c| \, \|f\|_p$ (homogeneity);
- $\|f+g\|_p \le \|f\|_p + \|g\|_p$ (Minkowski's inequality/the triangle inequality);
- $\|f\|_p = 0 \iff f = 0$ a.e.

Identifying functions that agree μ-almost everywhere (working with equivalence classes $[f]$ rather than individual functions f) turns the third property into the genuine positive-definiteness required of a norm, and $L^p(\Omega, \mu)$ becomes a normed vector space. The next goal is to show that this normed space is *complete*, i.e. a Banach space.

§10. 24 September 2026

We continue to work on a fixed measure space (Ω, Σ, μ).

Definition 10.1 (Equivalence class)**.** For a measurable function f, let

$$[f] = \{g \text{ measurable } : \ g = f \ \ \mu\text{-a.e.}\}. \tag{10.1}$$

Two functions with $g \in [f]$ are said to be *identified*: they represent the same element of L^p.

Remark 10.1 (The norm and operations are well defined on classes)**.** Since altering a function on a null set does not change the value of its integral, $\|g\|_p = \|f\|_p$ whenever $g \in [f]$; so $\|[f]\|_p := \|f\|_p$ is unambiguous. Likewise

$$[f] + [g] := [f + g], \qquad c\,[f] := [cf] \tag{10.2}$$

are well defined, independent of the representatives chosen: changing f or g on a set of measure zero changes $f + g$ only on a set of measure zero (a finite union of null sets is null), and similarly for scalar multiples.

Abuse of Notation. From now on we abuse notation and write f for the equivalence class $[f]$, rather than carrying the brackets throughout. Thus

$$L^p(\Omega, \mu) = \Big\{ f \text{ (an equivalence class) } : \ \|f\|_p < \infty \Big\}. \tag{10.3}$$

Proposition 10.1

With this identification, $L^p(\Omega, \mu)$ is a normed vector space: $\|\cdot\|_p$ satisfies homogeneity, the triangle inequality (Minkowski), and

$$\|f\|_p = 0 \iff f = 0 \text{ in } L^p, \tag{10.4}$$

where "$f = 0$ in L^p" now means literally f is the zero equivalence class, i.e. any representative of f vanishes μ-a.e.

Remark 10.2. Before this identification, $\|\cdot\|_p$ was only a *seminorm* on the space of functions (it could vanish on a nonzero function). Passing to equivalence classes is exactly what upgrades it to a genuine norm.

§10.i. The Riesz–Fischer Theorem

Definition 10.2 (Banach space)**.** A normed vector space is a **Banach space** if it is *complete*:

every Cauchy sequence converges to an element of the space. A sequence (f_n) in a normed space is *Cauchy* if for every $\varepsilon > 0$ there is N such that $\|f_m - f_n\| < \varepsilon$ whenever $m, n \geq N$.

Example 10.1

$\mathbb{Q}$ (with the usual absolute value) is not complete: a Cauchy sequence of rationals may converge to an irrational limit, which lies outside $\mathbb{Q}$. By contrast, $\mathbb{R}$ is complete: every Cauchy sequence of reals converges to a real number.

Theorem 10.2 (Riesz–Fischer)

For $1 \leq p \leq \infty$, $L^p(\Omega, \mu)$ is a Banach space.

Proof. Having already shown L^p is a normed vector space, it remains only to prove completeness. We treat $1 \leq p < \infty$ first, then $p = \infty$ separately. Let (f_n) be a Cauchy sequence in $L^p(\Omega, \mu)$: for every $\varepsilon > 0$ there is $N(\varepsilon)$ such that $\|f_m - f_n\|_p < \varepsilon$ whenever $m, n \geq N(\varepsilon)$.

Case 1: $1 \leq p < \infty$**.** For each $k = 1, 2, 3, \ldots$, apply the Cauchy property with $\varepsilon = 2^{-k}$ to choose an index $N(2^{-k})$, and select a subsequence $g_k := f_{n_k}$ (with n_k chosen increasing and, at each stage, at least as large as $N(2^{-k})$ and larger than the previously chosen index) so that

$$\|g_{k+1} - g_k\|_p \;\leq\; 2^{-k} \qquad \text{for every } k \geq 1. \tag{10.5}$$

(This is possible because the tail of the original sequence is eventually within 2^{-k} of any of its later terms, for each k.) Define

$$h_n(x) = |g_1(x)| + \sum_{k=1}^{n} |g_{k+1}(x) - g_k(x)|\,, \qquad n \geq 1. \tag{10.6}$$

Each $h_n \geq 0$, and (h_n) is pointwise increasing in n. By the Minkowski's inequality in L^p,

$$\|h_n\|_p \;\leq\; \|g_1\|_p + \sum_{k=1}^{n} \|g_{k+1} - g_k\|_p \;\leq\; \|g_1\|_p + \sum_{k=1}^{\infty} 2^{-k} = \|g_1\|_p + 1 =: M < \infty, \tag{10.7}$$

a bound independent of n. Let $h(x) = \lim_{n\to\infty} h_n(x) \in [0, \infty]$ (the limit exists in $[0, \infty]$ since (h_n) is increasing). By the Monotone Convergence Theorem applied to the monotone increasing sequence $h_n^p \to h^p$,

$$\int_\Omega h^p \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega h_n^p \, \mathrm{d}\mu = \lim_{n\to\infty} \|h_n\|_p^p \;\leq\; M^p < \infty, \tag{10.8}$$

so $h \in L^p(\Omega, \mu)$; in particular $h(x) < \infty$ for μ-a.e. x (Proposition 3.3). For a.e. x, $h(x) < \infty$, i.e. the series

$$\sum_{k=1}^{\infty} |g_{k+1}(x) - g_k(x)| \tag{10.9}$$

converges (absolutely) as a numerical series. Therefore, it converges without absolute value as well! Since

$$g_n(x) = g_1(x) + \sum_{k=1}^{n-1} (g_{k+1}(x) - g_k(x)) \tag{10.10}$$

is a telescoping sum, the sequence $(g_n(x))$ converges, for μ-a.e. x. Define

$$f(x) = \lim_{n\to\infty} g_n(x) \tag{10.11}$$

wherever this limit exists (a.e.), and $f(x) = 0$ on the remaining null set; this f is our candidate limit. For every n,

$$|g_n(x)| \le h_n(x) \le h(x) \qquad \text{for a.e. } x, \tag{10.12}$$

so letting $n \to \infty$, $|f(x)| \le h(x)$ a.e. as well. Since $h \in L^p$, this shows $f \in L^p(\Omega, \mu)$. Moreover

$$|f(x) - g_n(x)| \le |f(x)| + |g_n(x)| \le 2h(x) \qquad \text{a.e.,} \tag{10.13}$$

and $(2h)^p \in L^1$, while $|f - g_n|^p \to 0$ pointwise a.e. Therefore, Dominated Convergence Theorem applies to $|f - g_n|^p$ (dominated by $(2h)^p \in L^1$), giving

$$\lim_{n\to\infty} \int_\Omega |f - g_n|^p \, \mathrm{d}\mu = 0 \implies \lim_{n\to\infty} \|f - g_n\|_p = 0. \tag{10.14}$$

We have shown that the *subsequence* $g_n = f_{n_k}$ converges to f in L^p.

Claim 10.1: If $(x_n)_n$ is a Cauchy sequence in a metric space (X, d), and it has a subsequence $(x_{n_k})_k$ which converges to $x \in X$, then $x_n \to x$.

Proof. Fix $\varepsilon > 0$. There exists N_1 such that

$$d(x_{n_k}, x) \le \frac{\varepsilon}{2} \qquad \text{for } n_k \ge N_1. \tag{10.15}$$

There exists N_2 such that

$$d(x_n, x_m) \le \frac{\varepsilon}{2} \qquad \text{for } n, m \ge N_2. \tag{10.16}$$

Now, for all $n, n_k \ge \max\{N_1, N_2\}$,

$$d(x_n, x) \le d(x_n, x_{n_k}) + d(x_{n_k}, x) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{10.17}$$

Therefore, $x_n \to x$. □

Since f_n is a Cauchy sequence, and the subsequence g_n converges to f, therefore, $f_n \to f$ in L^p.

Case 2: $p = \infty$. Let (f_n) be Cauchy in $L^\infty(\Omega, \mu)$. For each pair (m, n) and each k, the inequality $|f_m(x) - f_n(x)| \le \|f_m - f_n\|_\infty$ and $|f_k(x)| \le \|f_k\|_\infty$ holds for a.e. x by definition of the essential supremum. Let Z be the (countable) union, over all pairs (m, n) and all k, of the

corresponding null exceptional sets. Then $\mu(Z) = 0$, and on $\Omega - Z$,

$$|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_\infty . \tag{10.18}$$

Given $\varepsilon > 0$, Cauchy-ness of (f_n) in L^∞ gives $N(\varepsilon)$ with $\|f_m - f_n\|_\infty < \varepsilon$ for $m, n \geq N(\varepsilon)$. Combined with (10.18), for every $x \in \Omega - Z$, and $m, n \geq N(\varepsilon)$,

$$|f_m(x) - f_n(x)| \leq \|f_m - f_n\|_\infty < \varepsilon. \tag{10.19}$$

So for each fixed $x \in \Omega - Z$, $(f_n(x))$ is a Cauchy sequence of real (or complex) numbers, hence converges. So we define

$$f(x) = \lim_{n\to\infty} f_n(x), \qquad x \in \Omega - Z, \tag{10.20}$$

and $f(x) = 0$ on Z. Letting $n \to \infty$ in $|f_m(x) - f_n(x)| < \varepsilon$ (for fixed $m \geq N(\varepsilon)$, $x \in \Omega - Z$) gives

$$|f_m(x) - f(x)| \leq \varepsilon, \tag{10.21}$$

for $x \in \Omega - Z$, $m \geq N(\varepsilon)$. i.e. for $m \geq N(\varepsilon)$,

$$\|f_m - f\|_\infty \leq \varepsilon. \tag{10.22}$$

This shows that $f_m \to f$ in L^∞. Now, taking $\varepsilon = 1$ and for some $m \geq N(1)$,

$$|f(x)| \leq |f(x) - f_m(x)| + |f_m(x)| \leq 1 + |f_m(x)| \leq 1 + \|f_m\|_\infty < \infty, \tag{10.23}$$

which is true a.e. Therefore, $f \in L^\infty$. ■

Corollary 10.3

If $f_n \to f$ in $L^p(\Omega, \mu)$ $(1 \leq p \leq \infty)$, then there is a subsequence (f_{n_k}) such that

$$f_{n_k}(x) \longrightarrow f(x) \qquad \text{pointwise for } \mu\text{-a.e. } x. \tag{10.24}$$

Remark 10.3 (This is *not* automatic)**.** Convergence $f_n \to f$ in the L^p norm is, a priori, a statement only about the size of $\|f_n - f\|_p$. It says nothing directly about the pointwise values $f_n(x)$ versus $f(x)$. (Indeed the various L^p norms are generally not comparable to one another without extra hypotheses, e.g. finiteness of the measure.) It is a genuine theorem, not a triviality, that norm convergence forces a subsequence to converge pointwise a.e.

For example, consider the sequence of functions $[0, 1] \to \mathbb{R}$ from Exercise 6.1:

$$f_1 = \chi_{[0,1]}$$
$$f_2 = 2\chi_{[0,\frac{1}{2}]},\ f_3 = 2\chi_{[\frac{1}{2},1]}$$
$$f_4 = 3\chi_{[0,\frac{1}{4}]},\ f_5 = 3\chi_{[\frac{1}{4},\frac{1}{2}]},\ f_6 = 3\chi_{[\frac{1}{2},\frac{3}{4}]},\ f_7 = 3\chi_{[\frac{3}{4},1]}$$
$$\dots$$

In other words, for $n = 2^k + m$, with k being the unique number satisfying $2^k \leq n < 2^{k+1}$, and $0 \leq m < 2^k$, we define

$$f_n = (k+1)\,\chi_{\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right]}. \tag{10.25}$$

Then

$$\|f_n - 0\|_1 = \int_0^1 |f_n(x)|\,\mathrm{d}x = \frac{k+1}{2^k} \to 0. \tag{10.26}$$

Now, given any $x \in [0,1]$, and for every $M > 0$, there are non-negative integers k, m such that

$$k > M \text{ and } m \leq 2^k x \leq m+1. \tag{10.27}$$

Now, $f_{2^k+m}(x) = k+1 > M$. Therefore, $\limsup_{k\to\infty} f_k(x) = \infty$. In particular, the pointwise limit $\lim_{k\to\infty} f_k(x)$ doesn't exist anywhere.

§10.ii. Density properties of L^p spaces

In order to do some of the exercises, we'll need to prove some density identities. To prove a property for all of L^p, it is often convenient to prove it first on a dense subclass and then extend.

Proposition 10.4

If $f \in L^p(\Omega, \mu)$ for $1 \leq p < \infty$, then there exists a sequence $\{\varphi_n\}$ of simple functions such that $\varphi_n \to f$ in L^p.

Proof. By Theorem 2.3 (applied on f^+ and f^- and then adding), there exists an increasing sequence of simple functions $\varphi_n \to f$ pointwise everywhere. Now,

$$|\varphi_n - f|^p \leq 2^p\,|f|^p \in L^1(\Omega, \mu). \tag{10.28}$$

Therefore, by Dominated Convergence Theorem,

$$\lim_{n\to\infty} \|\varphi_n - f\|_p^p = \lim_{n\to\infty} \int_\Omega |\varphi_n - f|^p\,\mathrm{d}\mu = \int_\Omega \lim_{n\to\infty} |\varphi_n - f|^p\,\mathrm{d}\mu = 0. \tag{10.29}$$

■

Note that since $f \in L^p$, if we write $\varphi_n = \sum_{i=1}^N a_i \chi_{E_i}$, then we must have $\mu(E_i) < \infty$ for all i. Because otherwise, $\int_\Omega \varphi_n^p\,\mathrm{d}\mu$ would be infinity. But then

$$\infty = \left|\int_\Omega \varphi_n^p\,\mathrm{d}\mu\right| \leq \int_\Omega |\varphi_n|^p\,\mathrm{d}\mu \leq \int_\Omega |f|^p\,\mathrm{d}\mu < \infty, \tag{10.30}$$

we get a contradiction!

Lemma 10.5

Suppose $1 \leq p < \infty$. Let $E \subseteq \mathbb{R}^d$ be measurable and $m(E) < \infty$. For every $\varepsilon > 0$, there is

a step function ψ such that $\|\psi - \chi_E\|_p \leq \varepsilon$.

Proof. We proceed similarly as Proposition 8.2. By Proposition 6.4(ii), choose a finite almost disjoint union of cubes

$$F' = \bigcup_{j=1}^{N} Q_j, \quad \text{with } m(E \triangle F') \leq \varepsilon^p. \tag{10.31}$$

Let $\psi := \sum_{j=1}^{N} \chi_{Q_j}$, a step function. For $x \in \bigcup_{j=1}^{N} Q_j^\circ$, since it's a disjoint union

$$\psi(x) = 1 = \chi_{F'}(x). \tag{10.32}$$

For $x \notin F$, $x \notin Q_j$ for any j, so

$$\psi(x) = 0 = \chi_{F'}(x). \tag{10.33}$$

Therefore, $\psi = \chi_{F'}$ on B^c, where $B = \bigcup_{j=1}^{N} (\partial Q_j)$ is the union of the boundaries of Q_j. Hence,

$$\Big\{x \in \mathbb{R}^d : \chi_E(x) \neq \psi(x)\Big\} \cap B^c = \Big\{x \in \mathbb{R}^d : \chi_E(x) \neq \chi_{F'}(x)\Big\} \cap B^c = \big(E \triangle F'\big) \cap B^c. \tag{10.34}$$

Now, on B^c, the points where $\chi_E \neq \psi$ are precisely $(E \triangle F') \cap B^c$. On B^c, ψ is either 0 or 1. Hence, $|\chi_E - \psi| \leq 1$ here. Therefore,

$$\begin{aligned} \|\psi - \chi_E\|_p^p &= \int_{\mathbb{R}} |\psi - \chi_E|^p \, \mathrm{d}m = \int_{B^c} |\psi - \chi_E|^p \, \mathrm{d}m \\ &= \int_{(E \triangle F') \cap B^c} |\psi - \chi_E|^p \, \mathrm{d}m \\ &\leq \int_{(E \triangle F') \cap B^c} 1 \, \mathrm{d}m \\ &= m\left(\big(E \triangle F'\big) \cap B^c\right) \leq \varepsilon^p. \end{aligned} \tag{10.35}$$

■

Lemma 10.6

Suppose $1 \leq p < \infty$. If $\varphi = \sum_{i=1}^{N} a_i \chi_{E_i}$ is a simple function in $\mathbb{R}^d$, with $m(E_i) < \infty$, then for every $\varepsilon > 0$ there exists a step function ψ such that $\|\varphi - \psi\|_p \leq \varepsilon$.

Proof. For each i, choose a step function ψ_i such that

$$\|\psi_i - \chi_{E_i}\|_p \leq \frac{\varepsilon}{N(1 + |a_i|)}. \tag{10.36}$$

Then we define the step function

$$\psi = \sum_{i=1}^{N} a_i \psi_i. \tag{10.37}$$

As a result,

$$\|\varphi-\psi\|_p=\left\|\sum_{i=1}^N a_i\chi_{E_i}-\sum_{i=1}^N a_i\psi_i\right\|_p\leq\sum_{i=1}^N|a_i|\,\|\chi_{E_i}-\psi_i\|_p$$
$$\leq\sum_{i=1}^N|a_i|\,\frac{\varepsilon}{N\left(1+|a_i|\right)}<\varepsilon. \tag{10.38}$$

■

Proposition 10.7

If $f\in L^p\left(\mathbb{R}^d\right)$ for $1\leq p<\infty$, given any $\varepsilon>0$, there exists a step function ψ such that $\|f-\psi\|_p\leq\varepsilon$.

Proof. By Proposition 10.4, we can find a sequence of simple functions φ_n such that $\|\varphi_n-f\|_p\to 0$. So we choose N large enough so that

$$\|f-\varphi_N\|_p\leq\frac{\varepsilon}{2}. \tag{10.39}$$

If we write $\varphi_n=\sum_{i=1}^N a_i\chi_{E_i}$, then we must have $m\left(E_i\right)<\infty$ for all i, according to the remark after Proposition 10.4. By Lemma 10.6, we choose a step function ψ such that

$$\|\varphi_N-\psi\|_p\leq\frac{\varepsilon}{2}. \tag{10.40}$$

Therefore,

$$\|f-\psi\|_p\leq\|f-\varphi_N\|_p+\|\varphi_N-\psi\|_p\leq\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon. \tag{10.41}$$

■

Corollary 10.8

If $f\in L^p\left(\mathbb{R}^d\right)$ for $1\leq p<\infty$, then there exists a sequence $\{\psi_n\}$ of step functions such that $\psi_n\to f$ in L^p.

Proof. Choose ψ_n with $\|f-\psi_n\|_p\leq\frac{1}{n}$. ■

Lemma 10.9

Suppose $1\leq p<\infty$. Let Q be a cube in $\mathbb{R}^d$. For every $\varepsilon>0$, there exists a compactly-supported continuous function g in $\mathbb{R}^d$ such that $\|\chi_Q-g\|_p\leq\varepsilon$.

Proof. Suppose $Q = \prod_{i=1}^{d} [a_i, b_i]$. Enlarge the cube in all sides to form the cube

$$Q' = \prod_{i=1}^{d} [a_i - \delta, b_i + \delta], \tag{10.42}$$

in such a way that

$$|Q'| \leq |Q| + \varepsilon^p. \tag{10.43}$$

Now, we define the function $g_i : \mathbb{R} \to \mathbb{R}$ as

$$g_i(x) = \begin{cases} 1 & \text{if } x \in [a, b], \\ 0 & \text{if } x \notin [a_i - \delta, b_i + \delta], \\ \text{linear} & \text{if } x \in (a_i - \delta, a_i) \text{ or } x \in [b_i, b_i + \delta]. \end{cases} \tag{10.44}$$

Then g_i is continuous. We define $g : \mathbb{R}^d \to \mathbb{R}$ as

$$g(\mathbf{x}) = \prod_{i=1}^{d} g_i(x_i). \tag{10.45}$$

Then g is conitinuous (as product of continuous functions), with $\operatorname{supp} g = Q'$ which is compact. Then g agrees with χ_Q on Q, and on $\mathbb{R}^d - Q'$. So they disagree only on $Q' - Q$, and on this region, $|\chi_Q - g| \leq 1$.

$$\|\chi_Q - g\|_p^p \int_{\mathbb{R}^d} |\chi_Q - g|^p \, \mathrm{d}m = \int_{Q'-Q} |\chi_Q - g|^p \, \mathrm{d}m \leq \int_{Q'-Q} 1 \, \mathrm{d}m = m(Q' - Q) \leq \varepsilon^p. \tag{10.46}$$

■

Lemma 10.10

Suppose $1 \leq p < \infty$. Let $\psi = \sum_{i=1}^{N} a_i \chi_{Q_i}$ be a step function on $\mathbb{R}^d$. For every $\varepsilon > 0$, there exists a compactly-supported continuous function g in $\mathbb{R}^d$ such that $\|\psi - g\|_p \leq \varepsilon$.

Proof. For each i, choose a compactly-supported continuous function g_i such that

$$\|g_i - \chi_{Q_i}\|_p \leq \frac{\varepsilon}{N(1 + |a_i|)}. \tag{10.47}$$

Then we define the compactly-supported continuous function

$$g = \sum_{i=1}^{N} a_i g_i. \tag{10.48}$$

As a result,

$$\|\psi - g\|_p = \left\| \sum_{i=1}^{N} a_i \chi_{Q_i} - \sum_{i=1}^{N} a_i g_i \right\|_p \leq \sum_{i=1}^{N} |a_i| \, \|\chi_{Q_i} - g_i\|_p \leq \sum_{i=1}^{N} |a_i| \frac{\varepsilon}{N(1 + |a_i|)} < \varepsilon. \tag{10.49}$$

■

Proposition 10.11

If $f \in L^p\left(\mathbb{R}^d\right)$ for $1 \leq p < \infty$, given any $\varepsilon > 0$, there exists a compactly-supported continuous function g such that $\|f - g\|_p \leq \varepsilon$.

Proof. By Proposition 10.7, we can find a step function ψ such that

$$\|f - \psi\|_p \leq \frac{\varepsilon}{2}. \tag{10.50}$$

By Lemma 10.10 we choose a compactly-supported continuous function g such that

$$\|\psi - g\|_p \leq \frac{\varepsilon}{2}. \tag{10.51}$$

Therefore,

$$\|f - g\|_p \leq \|f - \psi\|_p + \|\psi - g\|_p \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{10.52}$$

■

Corollary 10.12

If $f \in L^p\left(\mathbb{R}^d\right)$ for $1 \leq p < \infty$, then there exists a sequence $\{g_n\}$ of compactly-supported continuous functions such that $g_n \to f$ in L^p.

Proof. Choose g_n with $\|f - g_n\|_p \leq \frac{1}{n}$. ■

Therefore, the following classes are dense in L^p for $1 \leq p < \infty$:

(a) simple functions;

(b) step functions;

(c) compactly-supported continuous functions;

In fact, we can even take g_i in (10.44) to be a smooth bump function which gives 1 at $[a_i, b_i]$, and whose support is $[a_i - \delta_i, b_i - \delta_i]$. Thus compactly-supported smooth functions are also dense in L^p for $1 \leq p < \infty$. But $p = \infty$ is a different scenario altogether.

Remark 10.4. Let $f \in L^\infty(\Omega, \mu)$. Then $|f|$ is bounded almost everywhere. Note that by the construction of Theorem 2.3, for every non-negative function $f \in L^\infty$, we can construct a sequence of simple functions φ_n such that

$$0 \leq f(x) - \varphi_n(x) \leq \frac{1}{2^n} \quad \text{a.e.} \tag{10.53}$$

Hence,

$$\|f - \varphi_n\|_\infty \leq \frac{1}{2^n} \to 0. \tag{10.54}$$

For real functions,

$$\left\|\left(f^{+}-f^{-}\right)-\left(\varphi_{n}^{+}-\varphi_{n}^{-}\right)\right\|_{\infty} \leq\left\|f^{+}-\varphi_{n}^{+}\right\|_{\infty}+\left\|f^{+}-\varphi_{n}^{+}\right\|_{\infty} \rightarrow 0. \tag{10.55}$$

For complex functions,

$$\|(f+ig)-(\varphi_{n}+i\psi_{n})\| \leq\|f-\varphi_{n}\|+\|g-\psi_{n}\| \rightarrow 0. \tag{10.56}$$

Therefore, simple functions are dense in L^{∞}.

Remark 10.5. The same is not true step functions. For instance, if we choose

$$f(x)=\sin\left(\frac{1}{x}\right) \quad \text{for } x>0, \tag{10.57}$$

and $f(x)=0$ for $x \leq 0$. Then $|f| \leq 1$, everywhere, so that $f \in L^{\infty}(\mathbb{R})$. Choose any step function ψ. There exists some δ such that ψ is constant on $(0, \varepsilon)$. But since $\sin\left(\frac{1}{x}\right)$ is rapidly oscillating, it hits both $+1$ and -1 on $(0, \varepsilon)$. Therefore,

$$\|f-\psi\|_{\infty}=\operatorname{ess\,sup}|f-\psi| \geq \sup_{x \in(0, \varepsilon)}|f(x)-\psi(x)| \geq \max\{|1-c|,|-1-c|\} \geq 1. \tag{10.58}$$

Therefore, step functions are not dense in $L^{\infty}(\mathbb{R})$.

Remark 10.6. Continuous functions are not dense in $L^{\infty}(\mathbb{R})$. Consider the function

$$f(x)=\chi_{[0, \infty)} \in L^{\infty}(\mathbb{R}). \tag{10.59}$$

Assume for the sake of contradiction that there exists a continuous function g such that $\|f-g\|_{\infty}<\frac{1}{3}$. In other words, for almost all x,

$$f(x)-\frac{1}{3} \leq g(x) \leq f(x)+\frac{1}{3}. \tag{10.60}$$

Then $g(x) \leq \frac{1}{3}$ for almost all $x<0$, so by continuity,

$$g(0) \leq \frac{1}{3}. \tag{10.61}$$

On the other hand, $g(x) \geq \frac{2}{3}$ for almost all $x>0$, so by continuity,

$$g(0) \geq \frac{2}{3}, \tag{10.62}$$

which is a contradiction! Therefore, continuous functions are not dense in $L^{\infty}(\mathbb{R})$. As a result, the smaller class of compactly supported continuous functions also can't be dense in $L^{\infty}(\mathbb{R})$.

§10.iii. Exercises

Exercise 10.1. A function $\varphi : (a, b) \to \mathbb{R}$ is said to be convex if

$$\varphi\left((1-\lambda)x + \lambda y\right) \leq (1-\lambda)\varphi(x) + \lambda\varphi(y) \tag{10.63}$$

for all $x, y \in (a, b)$ and $0 \leq \lambda \leq 1$. If $\varphi : (a, b) \to \mathbb{R}$ is convex, prove that it is continuous.

Solution. Choose $x_0 \in (a, b)$. Then there exists R such that $[x_0 - R, x_0 + R] \subseteq (a, b)$. We shall prove that φ is Lipschitz on $\left[x_0 - \frac{R}{2}, x_0 + \frac{R}{2}\right]$, which will prove that φ is continuous at x_0.

First, we'll show that φ is bounded above on $[x_0 - R, x_0 + R]$. Indeed, any point in the inveral $[x_0 - R, x_0 + R]$ is a convex combination of its endpoints. Hence,

$$\varphi(x) \leq (1-\lambda)\varphi(x_0 - R) + \lambda\varphi(x_0 + R) \leq \max\left\{\varphi(x_0 - R), \varphi(x_0 + R)\right\} =: M. \tag{10.64}$$

Next, we'll show that φ is bounded below on $[x_0 - R, x_0 + R]$. Take $x \in [x_0 - R, x_0 + R]$. Then take $x' \in [x_0 - R, x_0 + R]$ such that x_0 is the midpoint of x and x'. Then

$$\varphi(x_0) \leq \frac{1}{2}\varphi(x) + \frac{1}{2}\varphi\left(x'\right) \implies \varphi(x) \geq 2\varphi(x_0) - \varphi\left(x'\right) \geq 2\varphi(x_0) - M. \tag{10.65}$$

Therefore, φ is bounded on $[x_0 - R, x_0 + R]$, say $|\varphi(x)| \leq K$. Now take $x, y \in \left[x_0 - \frac{R}{2}, x_0 + \frac{R}{2}\right]$, and WLOG $x < y$. Take $z = y + \frac{R}{2}$ and $w = x - \frac{R}{2}$. Then all $w, x, y, z \in [x_0 - R, x_0 + R]$ in that order.

We know that for a convex function $\varphi : (a, b) \to \mathbb{R}$ and $a < s < t < u < b$,

$$\frac{\varphi(t) - \varphi(s)}{t - s} \leq \frac{\varphi(u) - \varphi(t)}{u - t}. \tag{10.66}$$

Now, applying it to $w < x < y$ and $x < y < z$, we get

$$\frac{\varphi(x) - \varphi(w)}{R/2} \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \frac{\varphi(z) - \varphi(y)}{R/2}. \tag{10.67}$$

Now,

$$\frac{\varphi(z) - \varphi(y)}{R/2} \leq \frac{K + K}{R/2} = \frac{4K}{R} \text{ and } \frac{\varphi(x) - \varphi(w)}{R/2} \geq \frac{-K - K}{R/2} = -\frac{4K}{R}. \tag{10.68}$$

Therefore,

$$\left|\frac{\varphi(y) - \varphi(x)}{y - x}\right| \leq \frac{4K}{R}, \tag{10.69}$$

a constant that doesn't depend on x, y. Therefore, φ is Lipschitz on $\left[x_0 - \frac{R}{2}, x_0 + \frac{R}{2}\right]$. ■

Exercise 10.2. Determine the values of $a \in \mathbb{R}$ for which the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) = \frac{1}{|x|^a \left(1 + |x|\right)^{2a}} \tag{10.70}$$

belongs to $L^1(\mathbb{R})$.

Solution. $g \in L^1(\mathbb{R})$ if and only if $\int_{\mathbb{R}} g < \infty$.

$$\int_{-\infty}^{\infty} \frac{1}{|x|^a (1+|x|)^{2a}} \, \mathrm{d}x = 2\int_0^{\infty} \frac{1}{x^a (1+x)^{2a}} \, \mathrm{d}x \tag{10.71}$$

$$= 2\int_0^1 \frac{1}{x^a (1+x)^{2a}} \, \mathrm{d}x + 2\int_1^{\infty} \frac{1}{x^a (1+x)^{2a}} \, \mathrm{d}x. \tag{10.72}$$

So $g \in L^1(\mathbb{R})$ if and only if both these integrals are finite. Now, on $(0, 1]$,

$$\begin{aligned} 1 \le 1+x \le 2 &\implies 1 \le (1+x)^{2a} \le 2^{2a} = 4^a \\ &\implies 4^{-a} \le \frac{1}{(1+x)^{2a}} \le 1 \\ &\implies \frac{4^{-a}}{x^a} \le \frac{1}{x^a (1+x)^{2a}} \le \frac{1}{x^a}. \end{aligned}$$

(There is an implicit assumption that $a \ge 0$. But that's not an issue since the bounds are gonna just get swapped for the negative case.) Therefore, $\int_0^1 \frac{1}{x^a(1+x)^{2a}} \, \mathrm{d}x < \infty$ if and only if $\int_0^1 \frac{1}{x^a} \, \mathrm{d}x$ is finite. Clearly it's infinite for $a = 1$. For $a \neq 1$,

$$\int_0^1 \frac{1}{x^a} \, \mathrm{d}x = \left[\frac{x^{-a+1}}{-a+1}\right]_0^1 = \frac{1}{1-a} - \frac{1}{1-a} \lim_{x \to 0^+} x^{1-a}. \tag{10.73}$$

If $1 - a < 0$, the limit blows up. If $1 - a > 0$, then this is finite. Therefore, $\int_0^1 g(x) \, \mathrm{d}x < \infty$ if and only if $a < 1$.

$$\frac{1}{x^a (1+x)^{2a}} = \frac{1}{x^a x^{2a} \left(\frac{1}{x}+1\right)^{2a}} = x^{-3a} \left(1 + \frac{1}{x}\right)^{-2a}. \tag{10.74}$$

On $[1, \infty)$,

$$1 \le 1 + \frac{1}{x} \le 2 \implies 4^{-a} \le \frac{1}{\left(1+\frac{1}{x}\right)^{2a}} \le 1 \implies 4^{-a} x^{-3a} \le \frac{x^{-3a}}{\left(1+\frac{1}{x}\right)^{2a}} \le x^{-3a}. \tag{10.75}$$

(Again, there is an implicit assumption that $a \ge 0$. But that's not an issue since the bounds are gonna just get swapped for the negative case.) Therefore, $\int_1^{\infty} g(x) \, \mathrm{d}x < \infty$ if and only if $\int_1^{\infty} x^{-3a} \, \mathrm{d}x$ is finite. Clearly it's infinite for $a = \frac{1}{3}$. For $a \neq \frac{1}{3}$,

$$\int_1^{\infty} x^{-3a} \, \mathrm{d}x = \left[\frac{x^{1-3a}}{1-3a}\right]_1^{\infty} = \frac{1}{1-3a} \lim_{x \to \infty} x^{1-3a} - \frac{1}{1-3a}. \tag{10.76}$$

This will be finite if and only if $1 - 3a < 0$, i.e. $a > \frac{1}{3}$. Therefore, $g \in L^1(\mathbb{R})$ if and only if $\frac{1}{3} < a < 1$. ■

Exercise 10.3. Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be positive numbers with $\sum_{j=1}^n \alpha_j = 1$. Let $x_1, x_2, \ldots, x_n$ be

non-negative real numbers. Show that

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \leq \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n. \tag{10.77}$$

Solution. Note that $f(x) = \log x$ is a concave function, i.e.

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda) f(x) + \lambda f(y). \tag{10.78}$$

This is true because $f'' < 0$. We'll first show that for a concave function f and positive reals $\alpha_1, \ldots, \alpha_n$ with $\sum_{j=1}^n \alpha_j = 1$,

$$f\left(\sum_{j=1}^n \alpha_j x_j\right) \geq \sum_{j=1}^n \alpha_j f(x_j). \tag{10.79}$$

This follows by induction. The base case $n = 1$ is trivial. Suppose it's true for n-variables. Now, for $\sum_{j=1}^{n+1} \alpha_j = 1$,

$$f\left(\sum_{j=1}^{n+1} \alpha_j x_j\right) = f\left((1-\alpha_{n+1}) \sum_{j=1}^n \frac{\alpha_j}{1-\alpha_{n+1}} x_j + \alpha_{n+1} x_{n+1}\right) \tag{10.80}$$

$$\geq (1-\alpha_{n+1}) f\left(\sum_{j=1}^n \frac{\alpha_j}{1-\alpha_{n+1}} x_j\right) + \alpha_{n+1} f(x_{n+1}) \tag{10.81}$$

By induction hypothesis, since $\sum_{j=1}^n \frac{\alpha_j}{1-\alpha_{n+1}} = 1$,

$$f\left(\sum_{j=1}^n \frac{\alpha_j}{1-\alpha_{n+1}} x_j\right) \geq \sum_{j=1}^n \frac{\alpha_j}{1-\alpha_{n+1}} f(x_j). \tag{10.82}$$

Combining (10.81) and (10.82), we are done!

Now that we have proved (10.79), we shall apply it on log.

$$\log\left(\sum_{j=1}^n \alpha_j x_j\right) \geq \sum_{j=1}^n \alpha_j \log(x_j) = \log\left(\prod_{j=1}^n x_j^{\alpha_j}\right). \tag{10.83}$$

By exponentiating both sides, we get

$$\sum_{j=1}^n \alpha_j x_j \geq \prod_{j=1}^n x_j^{\alpha_j}. \tag{10.84}$$

■

Exercise 10.4. Assume that $f \in L^\infty([0,1])$ (with respect to the Lebesgue measure). Show that

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty. \tag{10.85}$$

Solution. In fact, this works for any finite measure space (Ω, Σ, μ). If $\|f\|_\infty = 0$, then $f = 0$ a.e. So the statement is trivially true. So we assume $\|f\|_\infty > 0$. Since $f \in L^\infty$, $|f(x)| \leq \|f\|_\infty$ a.e. Therefore,

$$\|f\|_p = \left(\int_\Omega |f|^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} \leq \left(\int_\Omega \|f\|_\infty^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} = \mu(\Omega)^{\frac{1}{p}} \|f\|_\infty . \tag{10.86}$$

Taking $p \to \infty$, we have

$$\limsup_{p\to\infty} \|f\|_p \leq \|f\|_\infty . \tag{10.87}$$

Choose $0 < \varepsilon < \|f\|_\infty$. We define

$$E_\varepsilon = \{x \in \Omega : |f(x)| > \|f\|_\infty - \varepsilon\} . \tag{10.88}$$

By the definition of essential supremum, $\mu(E_\varepsilon) > 0$. Now,

$$\|f\|_p = \left(\int_\Omega |f|^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} \geq \left(\int_{E_\varepsilon} |f|^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} \tag{10.89}$$

$$\geq \left(\int_{E_\varepsilon} (\|f\|_\infty - \varepsilon)^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} = (\|f\|_\infty - \varepsilon)\, \mu(E_\varepsilon)^{\frac{1}{p}} . \tag{10.90}$$

Taking $p \to \infty$, we have

$$\liminf_{p\to\infty} \|f\|_p \geq \|f\|_\infty - \varepsilon. \tag{10.91}$$

Since $\varepsilon > 0$ is arbitrarily small, we can take $\varepsilon \to 0^+$.

$$\liminf_{p\to\infty} \|f\|_p \geq \|f\|_\infty . \tag{10.92}$$

Combining (10.87) and (10.92), we have the desired limit $\lim_{p\to\infty} \|f\|_p = \|f\|_\infty$. ■

Exercise 10.5. If $f \in L^p(\mathbb{R}^d)$ and $v \in \mathbb{R}^d$, we define the translation $\tau_v f(x) := f(x - v)$.

(i) If $1 \leq p < \infty$, show that translation is a continuous operator in $L^p(\mathbb{R}^d)$, that is, show that

$$\lim_{|v|\to 0} \|\tau_v f - f\|_p = 0. \tag{10.93}$$

(ii) Show that this is not true when $p = \infty$.

Solution. (i) Let's first consider the case when f is a compactly-supported continuous function, where $\operatorname{supp} f = K$ is compact. WLOG, assume $|v| \leq 1$. Then

$$[\tau_v f - f](x) = f(x - v) - f(x) , \tag{10.94}$$

has the support contained in $K' = \{x : d(x, K) \leq 1\}$. This is bounded, since if K is bounded in a ball of radius R, K' is bounded in a ball of radius $R + 1$. Furthermore, this is closed, as

$$K' = d(-, K)^{-1} [0, 1] , \tag{10.95}$$

and $d(-, K)$ is continuous. Now,

$$\|\tau_v f - f\|_p^p = \int_{\mathbb{R}^d} |\tau_v f - f|^p \, \mathrm{d}m = \int_{K'} |\tau_v f - f|^p \, \mathrm{d}m. \tag{10.96}$$

Since f is a continuous function on a compact set, it's uniformly continuous. Therefore, given ε, we can find $\delta > 0$ such that

$$|f(x - v) - f(x)| < \frac{\varepsilon}{m(K')^{\frac{1}{p}}} \text{ whenever } |v| < \delta. \tag{10.97}$$

Since this holds uniformly for all $x \in K'$,

$$\|\tau_v f - f\|_p^p = \int_{K'} |\tau_v f - f|^p \, \mathrm{d}m \leq \int_{K'} \frac{\varepsilon^p}{m(K')} \mathrm{d}m = \frac{\varepsilon^p}{m(K')} \, m\left(K'\right) = \varepsilon^p. \tag{10.98}$$

Hence, $\|\tau_v f - f\|_p \leq \varepsilon$ whenever $|v| < \delta$, proving that

$$\lim_{|v| \to 0} \|\tau_v f - f\|_p = 0, \quad \text{for } f \in C_c\left(\mathbb{R}^d\right). \tag{10.99}$$

Now, we get back to the general case when $f \in L^p\left(\mathbb{R}^d\right)$. Given $\varepsilon > 0$, we can find a compactly-supported continuous function g such that

$$\|f - g\|_p \leq \frac{\varepsilon}{3}. \tag{10.100}$$

Since (10.99) holds for g, we can also find $\delta > 0$ such that whenever $|v| < \delta$,

$$\|\tau_v g - g\|_p \leq \frac{\varepsilon}{3}. \tag{10.101}$$

Now, notice that translation preserves norm, i.e. given h,

$$\|\tau_v h\|_p = \|h\|_p, \tag{10.102}$$

which follows from the fact that integration is translation invariant. Therefore,

$$\begin{aligned} \|\tau_v f - f\|_p &\leq \|\tau_v f - \tau_v g\|_p + \|\tau_v g - g\|_p + \|g - f\|_p \\ &= \|\tau_v (f - g)\|_p + \|\tau_v g - g\|_p + \|g - f\|_p \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon, \end{aligned} \tag{10.103}$$

whenever $|v| < \delta$. Therefore,

$$\lim_{|v| \to 0} \|\tau_v f - f\|_p = 0, \quad \text{for } f \in L^p\left(\mathbb{R}^d\right). \tag{10.104}$$

(ii) Consider the function $f = \chi_{[0,\infty)}$. It's in $L^\infty(\mathbb{R})$, since its essential supremum is 1. For

any v, the translation $\tau_v f = \chi_{[v,\infty)}$. Then

$$\tau_v f - f = \chi_{[v,\infty)} - \chi_{[0,\infty)} = \chi_{[0,v)} \text{ or } \chi_{[v,0)}. \tag{10.105}$$

Both of these has essential supremum 1. Therefore, $\lim_{|v|\to 0} \|\tau_v f - f\|_\infty = 1 \neq 0$.

■

Exercise 10.6. (i) Does there exist a function f that belongs to $L^2([0,1])$ but does not belong to any other $L^p([0,1])$, for $1 \leq p \leq \infty$, $p \neq 2$? If yes, give an example. If not, justify.

(ii) Does there exist a function f that belongs to $L^2(\mathbb{R})$ but does not belong to any other $L^p(\mathbb{R})$, for $1 \leq p \leq \infty$, $p \neq 2$? If yes, give an example. If not, justify.

(iii) Does there exist a function f that belongs to $L^1(\mathbb{R})$ but does not belong to any other $L^p(I)$ for any open interval $I \subset \mathbb{R}$ and any $p > 1$? If yes, give an example. If not, justify.

Solution. (i) No. From Remark 9.9, since $m([0,1]) = 1$, any $1 \leq p < 2$ has the property that $\|f\|_p \leq \|f\|_2$, so if $\|f\|_2$ is finite, then so is $\|f\|_p$ for $1 \leq p < 2$.

However, we can construct a function that is in $L^2([0,1])$ but not in L^p for $p > 2$. For instance, we define

$$f(x) = \frac{1}{\sqrt{x} \cdot (1 - \log x)}. \tag{10.106}$$

Then

$$\int_0^1 |f(x)|^2 \, \mathrm{d}x = \int_0^1 \frac{1}{x \cdot (1 - \log x)^2} \, \mathrm{d}x. \tag{10.107}$$

After the change of variables $u = 1 - \log x$, we have

$$\int \frac{1}{x \cdot (1 - \log x)^2} \, \mathrm{d}x = \int -\frac{1}{u^2} \, \mathrm{d}u = \frac{1}{u} = \frac{1}{1 - \log x} + C. \tag{10.108}$$

Therefore,

$$\int_0^1 |f(x)|^2 \, \mathrm{d}x = \left[\frac{1}{1 - \log x}\right]_0^1 = 1. \tag{10.109}$$

So $f \in L^2$. Now, for $p > 2$,

$$\int_0^1 |f(x)|^p \, \mathrm{d}x = \int_0^1 \frac{1}{x^{\frac{p}{2}} \cdot (1 - \log x)^p} \, \mathrm{d}x. = \int_\infty^0 \frac{e^{\frac{up}{2}}}{(1+u)^p} \left(-e^{-u} \, \mathrm{d}u\right), \tag{10.110}$$

for $u = -\log x$, so that $x = e^{-u}$, and $\mathrm{d}x = -e^{-u} \, \mathrm{d}u$. After a bit of simplification, we get

$$\int_0^1 |f(x)|^p \, \mathrm{d}x = \int_0^\infty \frac{e^{u\left(\frac{p}{2}-1\right)}}{(1+u)^p} \, \mathrm{d}u. \tag{10.111}$$

For $p > 2$, $e^{u\left(\frac{p}{2}-1\right)}$ grows exponentially while $(1+u)^p$ just grows polynomially, i.e.

$$\lim_{u\to\infty} \frac{e^{u\left(\frac{p}{2}-1\right)}}{(1+u)^p} = \infty, \tag{10.112}$$

proving that $\int_0^1 |f(x)|^p \, \mathrm{d}x = \infty$.

(ii) Yes. In fact, we'll prove that the following function $f : \mathbb{R} \to \mathbb{R}$ is in $L^p(\mathbb{R})$, but not in any $L^q(\mathbb{R})$ for $q \neq p$.

$$f(x) = \frac{1}{x^{\frac{1}{p}}\left(1+(\log x)^2\right)} \text{ for } x > 0, \tag{10.113}$$

and we can define $f(x) = 0$ for $x \leq 0$. Now,

$$\int_{\mathbb{R}} |f|^p \, \mathrm{d}m = \int_0^\infty \frac{1}{x\left(1+(\log x)^2\right)^p} \, \mathrm{d}x. \tag{10.114}$$

With the change of variables $u = \log x$, we have

$$\int_{\mathbb{R}} |f|^p \, \mathrm{d}m = \int_{-\infty}^\infty \frac{1}{(1+u^2)^p} \, \mathrm{d}u = 2\int_0^\infty \frac{1}{(1+u^2)^p} \, \mathrm{d}u \tag{10.115}$$

$$= 2\int_0^1 \frac{1}{(1+u^2)^p} \, \mathrm{d}u + 2\int_1^\infty \frac{1}{(1+u^2)^p} \, \mathrm{d}u. \tag{10.116}$$

On $[0,1]$, $\frac{1}{1+u^2} \leq 1$, so the integral is finite on $[0,1]$. Now, on $[1,\infty)$, $\frac{1}{1+u^2} \leq u^{-2}$. Hence,

$$\int_1^\infty \frac{1}{(1+u^2)^p} \, \mathrm{d}u \leq \int_1^\infty u^{-2p} \, \mathrm{d}u = \left[\frac{u^{1-2p}}{1-2p}\right]_1^\infty = \frac{1}{2p-1}, \tag{10.117}$$

since $1-2p < 0$, and hence $\lim_{u\to\infty} u^{1-2p} = 0$. Therefore, $f \in L^p(\mathbb{R})$. Now, we'll show that $f \notin L^q(\mathbb{R})$ for $q \neq p$.

$$\int_{\mathbb{R}} |f|^q \, \mathrm{d}m = \int_0^\infty \frac{1}{x^{\frac{q}{p}}\left(1+(\log x)^2\right)^q} \, \mathrm{d}x. \tag{10.118}$$

If $q > p$,

$$\int_{\mathbb{R}} |f|^q \, \mathrm{d}m \geq \int_0^{\frac{1}{e}} \frac{1}{x^{\frac{q}{p}}\left(1+(\log x)^2\right)^q} \, \mathrm{d}x = \int_\infty^1 \frac{e^{\frac{uq}{p}}}{(1+u^2)^q} \left(-e^{-u} \, \mathrm{d}u\right), \tag{10.119}$$

for $u = -\log x$, so that $x = e^{-u}$, and $\mathrm{d}x = -e^{-u} \, \mathrm{d}u$. After a bit of simplification, we get

$$\int_{\mathbb{R}} |f|^q \, \mathrm{d}m \geq \int_1^\infty \frac{e^{u\left(\frac{q}{p}-1\right)}}{(1+u^2)^q} \, \mathrm{d}u. \tag{10.120}$$

Since $q > p$, $\frac{q}{p} - 1 > 0$. So $e^{u\left(\frac{q}{p}-1\right)}$ grows exponentially while $\left(1+u^2\right)^q$ just grows polynomially, i.e.

$$\lim_{u\to\infty} \frac{e^{u\left(\frac{q}{p}-1\right)}}{(1+u^2)^q} = \infty, \tag{10.121}$$

proving that $\int_{\mathbb{R}} |f|^q \, \mathrm{d}m = \infty$. Now, consider the case $q < p$.

$$\int_{\mathbb{R}} |f|^q \, \mathrm{d}m \geq \int_e^{\infty} \frac{1}{x^{\frac{q}{p}} \left(1 + (\log x)^2\right)^q} \, \mathrm{d}x = \int_1^{\infty} \frac{e^{-\frac{uq}{p}}}{(1+u^2)^q} \left(e^u \, \mathrm{d}u\right), \tag{10.122}$$

for $u = \log x$, so that $x = e^u$, and $\mathrm{d}x = e^{-u} \, \mathrm{d}u$. After a bit of simplification, we get

$$\int_{\mathbb{R}} |f|^q \, \mathrm{d}m \geq \int_1^{\infty} \frac{e^{u\left(1-\frac{q}{p}\right)}}{(1+u^2)^q} \, \mathrm{d}u. \tag{10.123}$$

Since $q < p$, $1 - \frac{q}{p} > 0$. So $e^{u\left(1-\frac{q}{p}\right)}$ grows exponentially while $\left(1+u^2\right)^q$ just grows polynomially, i.e.

$$\lim_{u \to \infty} \frac{e^{u\left(1-\frac{q}{p}\right)}}{(1+u^2)^q} = \infty, \tag{10.124}$$

proving that $\int_{\mathbb{R}} |f|^q \, \mathrm{d}m = \infty$.

(iii) Yes. First, consider the function

$$g(x) = \frac{1}{|x| \left(\log |x|\right)^2}, \quad \text{for } 0 < |x| < \frac{1}{2}, \tag{10.125}$$

and $g(x) = 0$ elsewhere. First, note that $g \in L^1(\mathbb{R})$:

$$\int_{\mathbb{R}} |g| \, \mathrm{d}m = 2 \int_0^{\frac{1}{2}} \frac{1}{x \left(\log x\right)^2} \, \mathrm{d}x. \tag{10.126}$$

Now, take $u = \log x$, so that $\mathrm{d}u = \frac{1}{x} \, \mathrm{d}x$.

$$\int \frac{1}{x \left(\log x\right)^2} \, \mathrm{d}x = \int \frac{1}{u^2} \, \mathrm{d}u = -\frac{1}{u} = -\frac{1}{\log x} + C. \tag{10.127}$$

So,

$$\int_{\mathbb{R}} |g| \, \mathrm{d}m = 2 \left[-\frac{1}{\log x}\right]_0^{\frac{1}{2}} = \frac{2}{\log 2}. \tag{10.128}$$

Now we claim that $g \notin L^p((-\varepsilon, \varepsilon))$ for $0 < \varepsilon < \frac{1}{2}$ and $p > 1$.

$$\int_{-\varepsilon}^{\varepsilon} |g(x)|^p \, \mathrm{d}x = 2 \int_0^{\varepsilon} \frac{1}{x^p \left(-\log x\right)^{2p}} = 2 \int_{\infty}^{-\log \varepsilon} \frac{e^{up}}{u^{2p}} \left(-e^{-u} \, \mathrm{d}u\right) \tag{10.129}$$

for $u = -\log x$, so that $x = e^{-u}$, and $\mathrm{d}x = -e^{-u} \, \mathrm{d}u$. After a bit of simplification, we get

$$\int_{-\varepsilon}^{\varepsilon} |g(x)|^p \, \mathrm{d}x = \int_{-\log \varepsilon}^{\infty} \frac{e^{u(p-1)}}{u^{2p}} \, \mathrm{d}u. \tag{10.130}$$

Since $p - 1 > 0$, $e^{u(p-1)}$ grows exponentially while u^{2p} just grows polynomially, i.e.

$$\lim_{u \to \infty} \frac{e^{u(p-1)}}{u^{2p}} = \infty, \tag{10.131}$$

proving that $\int_{-\varepsilon}^{\varepsilon} |g(x)|^p \, \mathrm{d}x = \infty$, i.e. $g \notin L^p((-\varepsilon, \varepsilon))$.

Now, let $\{q_n\}_{n=1}^{\infty}$ be an enumeration of $\mathbb{Q}$. We define

$$f(x) = \sum_{n=1}^{\infty} \frac{1}{2^n} \, g(x - q_n). \tag{10.132}$$

By applying Proposition 2.8, we get

$$\int_{\mathbb{R}} |f| \, \mathrm{d}m \leq \sum_{n=1}^{\infty} \frac{1}{2^n} \int_{-\infty}^{\infty} |g(x - q_n)| \, \mathrm{d}x = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{2}{\log 2} = \frac{2}{\log 2}. \tag{10.133}$$

Therefore, $f \in L^1(\mathbb{R})$. Now, let $I \subseteq \mathbb{R}$ be an arbitrary open interval. It contains some rational number, say q_k. Then there exists some $\varepsilon > 0$ such that $(q_k - \varepsilon, q_k + \varepsilon) \subseteq I$. WLOG, $\varepsilon < \frac{1}{2}$. Now, on I,

$$f(x) \geq \frac{1}{2^k} \, g(x - q_k), \tag{10.134}$$

so we have

$$\int_I |f|^p \, \mathrm{d}m \geq \int_{q_k - \varepsilon}^{q_k + \varepsilon} \left| \frac{1}{2^k} \, g(x - q_k) \right|^p \mathrm{d}x = \frac{1}{2^{kp}} \int_{-\varepsilon}^{\varepsilon} |g(u)|^p \, \mathrm{d}u, \tag{10.135}$$

for $u = x - q_k$. This integral diverges, as we have just seen. Therefore, $f \notin L^p(I)$ for any $I \subseteq \mathbb{R}$ and $p > 1$.

■

Exercise 10.7. Let $g : \mathbb{R} \to \mathbb{C}$ be a bounded, measurable, periodic function with period 1, and let f be a function in $L^1(\mathbb{R})$. Show that

$$\lim_{n \to \infty} \int_{\mathbb{R}} f(x) g(nx) \, \mathrm{d}x = \left(\int_0^1 g(x) \, \mathrm{d}x \right) \left(\int_{\mathbb{R}} f(x) \, \mathrm{d}x \right). \tag{10.136}$$

Solution. Note that if (10.136) holds for f, then it holds for cf. Furthermore, if it holds for $f = f_1$ and $f = f_2$, then it holds for $f = f_1 + f_2$.

First, we'll prove that (10.136) holds for $\chi_{[a,b]}$. Indeed,

$$\int_{\mathbb{R}} \chi_{[a,b]} g(nx) \, \mathrm{d}x = \int_a^b g(nx) \, \mathrm{d}x = \frac{1}{n} \int_{na}^{nb} g(u) \, \mathrm{d}u. \tag{10.137}$$

Since g is periodic with period 1, its integral over any interval with length 1 is the same, $I := \int_0^1 g(x) \, \mathrm{d}x$. Therefore,

$$\int_{na}^{nb} g(u) \, \mathrm{d}u = I \times (\text{number of length-1 intervals in } [na, nb]) + \int_J g(x) \, \mathrm{d}x, \tag{10.138}$$

where J is an interval of length less than 1. The number of length-1 intervals in $[na, nb]$ is $\lfloor nb - na \rfloor$. Therefore, the RHS of (10.137) becomes

$$\frac{1}{n} \times I \times \lfloor n(b-a) \rfloor + \frac{1}{n} \int_J g\left(x\right)\, \mathrm{d}x. \tag{10.139}$$

The second part is bounded by

$$\left| \frac{1}{n} \int_J g\left(x\right)\, \mathrm{d}x \right| \leq \frac{1}{n} \int_{[0,1]} |g(x)|\, \mathrm{d}x \leq \frac{M}{n} \to 0, \tag{10.140}$$

where M is the bound of g, i.e. $|g| \leq M$. For the first part, note that $\lim_{n\to\infty} \frac{\lfloor nx \rfloor}{n} = x$ for any real x (this can be easily proven by squeeze theorem). Therefore,

$$\frac{1}{n} \times I \times \lfloor n(b-a) \rfloor \to I\left(b-a\right) = \left(\int_0^1 g(x)\, \mathrm{d}x \right) \left(\int_{\mathbb{R}} \chi_{[a,b]}(x)\, \mathrm{d}x \right). \tag{10.141}$$

Therefore, (10.136) holds for $\chi_{[a,b]}$. By linearity, it holds for any step function.

Now, let $f \in L^1(\mathbb{R})$ be arbitrary, and fix $\varepsilon > 0$. Choose a step function ψ such that

$$\left\| f - \psi \right\|_1 \leq \frac{\varepsilon}{3M}. \tag{10.142}$$

Now,

$$\begin{aligned}
&\left| \int_{\mathbb{R}} f(x) g(nx)\, \mathrm{d}x - \left(\int_0^1 g(x)\, \mathrm{d}x \right) \left(\int_{\mathbb{R}} f(x)\, \mathrm{d}x \right) \right| \\
&= \left| \int_{\mathbb{R}} \left[f(x) - \psi(x) \right] g(nx)\, \mathrm{d}x + \int_{\mathbb{R}} \psi(x) g(nx)\, \mathrm{d}x - I \int_{\mathbb{R}} \left[f(x) - \psi(x) \right]\, \mathrm{d}x - I \int_{\mathbb{R}} \psi(x)\, \mathrm{d}x \right| \\
&\leq \int_{\mathbb{R}} |f(x) - \psi(x)|\, |g(nx)|\, \mathrm{d}x + |I| \int_{\mathbb{R}} |f(x) - \psi(x)|\, \mathrm{d}x + \left| \int_{\mathbb{R}} \psi(x) g(nx)\, \mathrm{d}x - I \int_{\mathbb{R}} \psi(x)\, \mathrm{d}x \right|.
\end{aligned}$$

Since $\int_{\mathbb{R}} \psi(x) g(nx)\, \mathrm{d}x \to I \int_{\mathbb{R}} \psi(x)\, \mathrm{d}x$, we can find N such that for $n \geq N$,

$$\left| \int_{\mathbb{R}} \psi(x) g(nx)\, \mathrm{d}x - I \int_{\mathbb{R}} \psi(x)\, \mathrm{d}x \right| \leq \frac{\varepsilon}{3}. \tag{10.143}$$

Furthermore,

$$|I| = \left| \int_0^1 g(x)\, \mathrm{d}x \right| \leq \int_0^1 |g(x)|\, \mathrm{d}x \leq \int_0^1 M\, \mathrm{d}x = M. \tag{10.144}$$

Therefore, we have that for $n \geq N$,

$$\begin{aligned}
&\left| \int_{\mathbb{R}} f(x)g(nx)\,\mathrm{d}x - \left(\int_0^1 g(x)\,\mathrm{d}x \right) \left(\int_{\mathbb{R}} f(x)\,\mathrm{d}x \right) \right| \\
\leq\ & \int_{\mathbb{R}} |f(x) - \psi(x)|\,|g(nx)|\,\mathrm{d}x + M \int_{\mathbb{R}} |f(x) - \psi(x)|\,\mathrm{d}x + \frac{\varepsilon}{3} \\
\leq\ & M \int_{\mathbb{R}} |f(x) - \psi(x)|\,\mathrm{d}x + M \int_{\mathbb{R}} |f(x) - \psi(x)|\,\mathrm{d}x + \frac{\varepsilon}{3} \\
\leq\ & M \cdot \frac{\varepsilon}{3M} + M \cdot \frac{\varepsilon}{3M} + \frac{\varepsilon}{3} = \varepsilon.
\end{aligned} \tag{10.145}$$

Therefore, $\int_{\mathbb{R}} f(x)g(nx)\,\mathrm{d}x \to \left(\int_0^1 g(x)\,\mathrm{d}x \right) \left(\int_{\mathbb{R}} f(x)\,\mathrm{d}x \right)$. ■

Exercise 10.8. Let $f \in L^1(\mathbb{R}^d)$. For any $\varepsilon > 0$, show that there exists a $\delta > 0$ such that for every Lebesgue measurable set $E \subset \mathbb{R}^d$ with $m(E) < \delta$ we have

$$\left| \int_E f(x)\,\mathrm{d}x \right| < \varepsilon. \tag{10.146}$$

Solution. We define

$$f_n(x) = \begin{cases} f(x) & \text{if } |f(x)| \leq n, \\ 0 & \text{if } |f(x)| > n. \end{cases} \tag{10.147}$$

In other words, $f_n = f\chi_{\{|f| \leq n\}}$. Then we have $f - f_n = f\chi_{\{|f| > n\}}$. Since f is integrable, it's finite a.e. Therefore, $f_n \to f$ pointwise a.e. Now,

$$|f - f_n| = |f|\,\chi_{\{|f| > n\}} \leq |f| \quad \text{pointwise eveyrwhere .} \tag{10.148}$$

so $f - f_n$ is dominated by the integrable function $|f|$. By Dominated Convergence Theorem,

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |f - f_n|\,\mathrm{d}m = \int_{\mathbb{R}^d} \lim_{n\to\infty} |f - f_n|\,\mathrm{d}m = 0. \tag{10.149}$$

Given $\varepsilon > 0$, we pick N such that

$$\int_{\mathbb{R}^d} |f - f_N|\,\mathrm{d}m < \frac{\varepsilon}{2}. \tag{10.150}$$

We pick $\delta = \frac{\varepsilon}{2N}$. Now, given measurable E with $m(E) < \delta = \frac{\varepsilon}{2N}$,

$$\left| \int_E f\,\mathrm{d}m \right| \leq \int_E |f|\,\mathrm{d}m \leq \int_E |f - f_N|\,\mathrm{d}m + \int_E |f_N|\,\mathrm{d}m. \tag{10.151}$$

Now, since $|f_N(x)| \leq N$ for every x,

$$\int_E |f_N|\,\mathrm{d}m \leq \int_E N\,\mathrm{d}m = N \cdot m(E) = N\delta = \frac{\varepsilon}{2}; \tag{10.152}$$

$$\int_E |f - f_N|\,\mathrm{d}m \leq \int_{\mathbb{R}^d} |f - f_N|\,\mathrm{d}m < \frac{\varepsilon}{2}. \tag{10.153}$$

Plugging these into (10.151), we have

$$\left|\int_E f\,\mathrm{d}m\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{10.154}$$

■

Alternate Solution. In fact, this is true for any measure space (Ω, Σ, μ). Assume the contrary. Then there exists some $\varepsilon > 0$ such that for all $\delta > 0$, there is a measurable set A_δ such that

$$\mu(A_\delta) < \delta \text{ and } \left|\int_{A_\delta} f\,\mathrm{d}\mu\right| \geq \varepsilon. \tag{10.155}$$

Now, for $\delta = 2^{-n}$, we obtain the sets $A_{2^{-n}}$, and we define the function

$$g_n = \chi_{A_{2^{-n}}} f. \tag{10.156}$$

Then each g_n is dominated by $|f| \in L^1$. We claim that $g_n \to 0$ a.e. By Borel–Cantelli lemma, the set of points that are in infinitely many $A_{2^{-n}}$'s has measure 0, since

$$\sum_{n=1}^{\infty} \mu(A_{2^{-n}}) < \sum_{n=1}^{\infty} 2^{-n} < \infty. \tag{10.157}$$

Therefore, for a.e. $x \in \Omega$, they only belong to a finitely many $A_{2^{-n}}$'s, and hence the sequence $g_n(x)$ is eventually 0. Therefore, $g_n \to 0$ a.e. Now, by Dominated Convergence Theorem,

$$\varepsilon \leq \lim_{n\to\infty} \left|\int_{A_{2^{-n}}} f\,\mathrm{d}\mu\right| = \lim_{n\to\infty} \left|\int_\Omega g_n\,\mathrm{d}\mu\right| = \left|\int_\Omega \lim_{n\to\infty} g_n\,\mathrm{d}\mu\right| = \left|\int_\Omega 0\,\mathrm{d}\mu\right| = 0, \tag{10.158}$$

which is a contradiction! ■

Exercise 10.9. Let $\{f_n\} \subset L^1([0,1])$ with

(i) $f_n \to f$ a.e., $f \in L^1([0,1])$;

(ii) $\|f_n\|_1 \to \|f\|_1$.

Show that $f_n \to f$ in $L^1([0,1])$.

Solution. We're gonna prove the following general result:

> Let (Ω, Σ, μ) be a measure space, and $f_n \in L^1(\Omega, \mu)$ be a sequence of integrable functions that converge pointwise a.e. to $f \in L^1(\Omega, \mu)$. Then
>
> $$f_n \to f \text{ in } L^1(\Omega, \mu) \iff \int_\Omega |f_n|\,\mathrm{d}\mu \to \int_\Omega |f|\,\mathrm{d}\mu. \tag{10.159}$$

($\Rightarrow$) By reverse triangle inequality in L^1,

$$\left|\|f\|_1 - \|f_n\|_1\right| \leq \|f - f_n\|_1 . \tag{10.160}$$

Since the RHS converges to 0 (by hypothesis $f_n \to f$ in L^1), therefore, the LHS also converges to 0, i.e. $\int_\Omega |f_n| \, \mathrm{d}\mu \to \int_\Omega |f| \, \mathrm{d}\mu$.

($\Leftarrow$) We are given that $\int_\Omega |f_n| \, \mathrm{d}\mu \to \int_\Omega |f| \, \mathrm{d}\mu$. We need to show that $\int_\Omega |f - f_n| \, \mathrm{d}\mu \to 0$. Note that,

$$|f(x)| + |f_n(x)| - |f(x) - f_n(x)| \geq 0, \tag{10.161}$$

by triangle inequality. Now, applying Fatou's Lemma on $|f| + |f_n| - |f - f_n|$,

$$\int_\Omega \liminf_{n\to\infty} \left[|f| + |f_n| - |f - f_n|\right] \mathrm{d}\mu \leq \liminf_{n\to\infty} \int_\Omega \left[|f| + |f_n| - |f - f_n|\right] \mathrm{d}\mu. \tag{10.162}$$

$|f| + |f_n| - |f - f_n|$ converges pointwise a.e. to $2|f|$. So the LHS is $\int_\Omega 2|f| \, \mathrm{d}\mu$. In the RHS, applying linearity and $\int_\Omega |f_n| \, \mathrm{d}\mu \to \int_\Omega |f| \, \mathrm{d}\mu$, we get

$$\text{RHS} = \liminf_{n\to\infty} \left[\int_\Omega |f| \, \mathrm{d}\mu + \int_\Omega |f_n| \, \mathrm{d}\mu - \int_\Omega |f - f_n| \, \mathrm{d}\mu\right] \tag{10.163}$$

$$= 2\int_\Omega |f| \, \mathrm{d}\mu - \limsup_{n\to\infty} \int_\Omega |f - f_n| \, \mathrm{d}\mu. \tag{10.164}$$

Cancelling out $2\int_\Omega |f| \, \mathrm{d}\mu$ from LHS $\leq$ RHS, we get

$$\limsup_{n\to\infty} \int_\Omega |f - f_n| \, \mathrm{d}\mu \leq 0. \tag{10.165}$$

Since this is a nonnegative sequence with lim sup 0, therefore, the limit must be 0, i.e. $\|f - f_n\|_1 \to 0$. ■

Exercise 10.10. Let $B = \left\{f \in L^2(\mathbb{R}) : \|f\|_2 \leq 1\right\}$. Let $\{f_n\} \subset B$ be a sequence such that $f_n \to f$ a.e. in $\mathbb{R}$. Show that $f \in B$ and that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x) g(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} f(x) g(x) \, \mathrm{d}x \tag{10.166}$$

for all $g \in L^2(\mathbb{R})$.

Solution. The fact that $f \in B$ follows from Fatou's Lemma:

$$\|f\|_2^2 = \int_{\mathbb{R}} |f|^2 \, \mathrm{d}m \leq \liminf_{n\to\infty} \int_{\mathbb{R}} |f_n|^2 \, \mathrm{d}m \leq \liminf_{n\to\infty} 1 \leq 1. \tag{10.167}$$

Now, fix $\varepsilon > 0$. Since $\int_{\mathbb{R}} |g|^2 \, \mathrm{d}m$ is finite, we can choose M large enough so that

$$\int_{|x|>M} |g(x)|^2 \, \mathrm{d}x \leq \frac{\varepsilon^2}{36}. \tag{10.168}$$

Now,

$$\left|\int_{\mathbb{R}} f_n g \,\mathrm{d}m - \int_{\mathbb{R}} fg \,\mathrm{d}m\right| \leq \int_{\mathbb{R}} |(f_n - f)g| \,\mathrm{d}m$$
$$= \int_{[-M,M]} |(f_n - f)g| \,\mathrm{d}m + \int_{|x|>M} |(f_n - f)g| \,\mathrm{d}m. \tag{10.169}$$

By Hölder's inequality,

$$\begin{aligned}\int_{|x|>M} |(f_n - f)g| \,\mathrm{d}m &\leq \|f_n - f\|_{L^2(\mathbb{R}-[-M,M])} \|g\|_{L^2(\mathbb{R}-[-M,M])} \\ &\leq \|f_n - f\|_{L^2(\mathbb{R})} \|g\|_{L^2(\mathbb{R}-[-M,M])} \\ &\leq \left[\|f_n\|_{L^2(\mathbb{R})} + \|f\|_{L^2(\mathbb{R})}\right] \frac{\varepsilon}{6} \\ &\leq 2 \cdot \frac{\varepsilon}{6} = \frac{\varepsilon}{3}.\end{aligned} \tag{10.170}$$

Now, let's bound $\int_{[-M,M]} |(f_n - f)g| \,\mathrm{d}m$. By Exercise 10.8, since $|g|^2 \in L^1(\mathbb{R})$, we can find $\delta > 0$ such that for every measurable set E with $m(E) < \delta$,

$$\int_E \left|g^2\right| \,\mathrm{d}m \leq \frac{\varepsilon^2}{36}. \tag{10.171}$$

Since $f_n \to f$ pointwise a.e. by Egorov's theorem, there exists a measurable set $A \subseteq [-M, M]$ such that $m([-M, M] - A) < \delta$, and $f_n \to f$ uniformly on A.

$$\int_{[-M,M]} |(f_n - f)g| \,\mathrm{d}m = \int_A |(f_n - f)g| \,\mathrm{d}m + \int_{[-M,M]-A} |(f_n - f)g| \,\mathrm{d}m. \tag{10.172}$$

Again by Hölder's inequality,

$$\int_{[-M,M]-A} |(f_n - f)g| \,\mathrm{d}m \leq 2\left(\int_{[-M,M]-A} |g|^2 \,\mathrm{d}m\right)^{\frac{1}{2}} \leq \frac{\varepsilon}{3}. \tag{10.173}$$

Since $f_n \to f$ uniformly on A, we can choose N such that for every $n \geq N$,

$$\sup_{x\in A} |f_n - f| \leq \frac{\varepsilon}{3\sqrt{2M}\,\|g\|_2}. \tag{10.174}$$

Now,

$$\int_A |(f_n - f)g| \,\mathrm{d}m \leq \frac{\varepsilon}{3\sqrt{2M}\,\|g\|_2} \int_A |g| \,\mathrm{d}m. \tag{10.175}$$

Again by Hölder's inequality,

$$\int_A |g| \,\mathrm{d}m \leq \left(\int_A 1^2 \,\mathrm{d}m\right)^{\frac{1}{2}} \left(\int_A |g|^2 \,\mathrm{d}m\right)^{\frac{1}{2}} \leq \sqrt{m(A)}\,\|g\|_2 \leq \sqrt{2M}\,\|g\|_2. \tag{10.176}$$

Therefore,

$$\int_A |(f_n - f)g| \,\mathrm{d}m \leq \frac{\varepsilon}{3\sqrt{2M}\,\|g\|_2} \cdot \sqrt{2M}\,\|g\|_2 = \frac{\varepsilon}{3}. \tag{10.177}$$

Now, for $n \geq N$,

$$\begin{aligned}\left|\int_{\mathbb{R}} f_n g \, \mathrm{d}m - \int_{\mathbb{R}} fg \, \mathrm{d}m\right| &\leq \int_{[-M,M]} |(f_n - f)g| \, \mathrm{d}m + \int_{|x|>M} |(f_n - f)g| \, \mathrm{d}m \\ &= \int_A |(f_n - f)g| \, \mathrm{d}m + \int_{[-M,M]-A} |(f_n - f)g| \, \mathrm{d}m \\ &\quad + \int_{|x|>M} |(f_n - f)g| \, \mathrm{d}m \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \end{aligned} \tag{10.178}$$

■

Alternate Solution. We can also proceed similarly as Exercise 10.7. Note that if (10.166) holds for g, then it holds for cg. Furthermore, if it holds for $g = g_1$ and $g = g_2$, then it holds for $g = g_1 + g_2$.

First, we'll prove that (10.166) holds for $g = \chi_{[a,b]}$. Indeed,

$$\left|\int_{\mathbb{R}} f_n \chi_{[a,b]} \, \mathrm{d}m - \int_{\mathbb{R}} f\chi_{[a,b]} \, \mathrm{d}m\right| = \left|\int_{[a,b]} (f_n - f) \, \mathrm{d}m\right| \leq \int_{[a,b]} |f_n - f| \, \mathrm{d}m. \tag{10.179}$$

Fix $\varepsilon > 0$. Now, since $f_n \to f$ pointwise a.e. on $[a,b]$, there exists $A \subseteq [a,b]$ such that

$$f_n \to f \text{ uniformly on } [a,b] - A, \text{ and } m(A) \leq \frac{\varepsilon^2}{16}. \tag{10.180}$$

Now, choose N such that for $n \geq N$,

$$\sup_{x \in [a,b]-A} |f_n(x) - f(x)| \leq \frac{\varepsilon}{2(b-a)}. \tag{10.181}$$

Then for $n \geq N$,

$$\int_{[a,b]-A} |f_n - f| \, \mathrm{d}m \leq \int_{[a,b]-A} \frac{\varepsilon}{2(b-a)} \, \mathrm{d}m = \frac{\varepsilon}{2(b-a)} m([a,b] - A) \leq \frac{\varepsilon}{2}. \tag{10.182}$$

On A, by Hölder's inequality,

$$\int_A |f_n - f| \, \mathrm{d}m \leq \left(\int_A |f_n - f|^2 \, \mathrm{d}m\right)^{\frac{1}{2}} \left(\int_A 1^2 \, \mathrm{d}m\right)^{\frac{1}{2}} \tag{10.183}$$

$$\leq \|f_n - f\|_2 \sqrt{m(A)} \tag{10.184}$$

$$\leq (\|f_n\| + \|f\|) \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2}. \tag{10.185}$$

Combining (10.182) and (10.185), we have

$$\int_{[a,b]} |f_n - f| \, \mathrm{d}m \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \tag{10.186}$$

for $n \geq N$. Therefore, (10.166) holds for $g = \chi_{[a,b]}$. By linearity, it holds for any step function.

Now let $g \in L^2$ be arbitrary, and fix $\varepsilon > 0$. Choose a step function ψ such that

$$\|g - \psi\|_2 \leq \frac{\varepsilon}{3}. \tag{10.187}$$

Now,

$$\begin{aligned}
&\left| \int_{\mathbb{R}} f_n g \, \mathrm{d}m - \int_{\mathbb{R}} f g \, \mathrm{d}m \right| \\
&= \left| \int_{\mathbb{R}} f_n \left(g - \psi\right) \, \mathrm{d}m - \int_{\mathbb{R}} f \left(g - \psi\right) \, \mathrm{d}m + \int_{\mathbb{R}} f_n \psi \, \mathrm{d}m - \int_{\mathbb{R}} f \psi \, \mathrm{d}m \right| \\
&\leq \int_{\mathbb{R}} \left| f_n \left(g - \psi\right) \right| \, \mathrm{d}m + \int_{\mathbb{R}} \left| f \left(g - \psi\right) \right| \, \mathrm{d}m + \left| \int_{\mathbb{R}} f_n \psi \, \mathrm{d}m - \int_{\mathbb{R}} f \psi \, \mathrm{d}m \right|
\end{aligned}$$

Since $\int_{\mathbb{R}} f_n \psi \, \mathrm{d}m \to \int_{\mathbb{R}} f \psi \, \mathrm{d}m$, we can choose N such that for every $n \geq N$,

$$\left| \int_{\mathbb{R}} f_n \psi \, \mathrm{d}m - \int_{\mathbb{R}} f \psi \, \mathrm{d}m \right| \leq \frac{\varepsilon}{3}. \tag{10.188}$$

By Hölder's inequality,

$$\int_{\mathbb{R}} \left| f_n \left(g - \psi\right) \right| \, \mathrm{d}m \leq \|f_n\|_2 \, \|g - \psi\|_2 \leq \frac{\varepsilon}{3}. \tag{10.189}$$

Similarly,

$$\int_{\mathbb{R}} \left| f \left(g - \psi\right) \right| \, \mathrm{d}m \leq \frac{\varepsilon}{3}. \tag{10.190}$$

Therefore, for every $n \geq N$, we have

$$\left| \int_{\mathbb{R}} f_n g \, \mathrm{d}m - \int_{\mathbb{R}} f g \, \mathrm{d}m \right| \leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon. \tag{10.191}$$

Hence, $\int_{\mathbb{R}} f_n g \, \mathrm{d}m \to \int_{\mathbb{R}} f g \, \mathrm{d}m$. ■

Exercise 10.11. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lebesgue measurable function such that

$$m\{x : |f(x)| > \lambda\} \leq \frac{C}{\lambda^2} \tag{10.192}$$

for some fixed constant $C > 0$ and all $\lambda > 0$. Prove that there exists a constant C_1 such that for any Lebesgue measurable set $E \subset \mathbb{R}^d$ of finite and positive measure we have

$$\int_E |f(x)| \, \mathrm{d}x \leq C_1 \sqrt{m(E)}. \tag{10.193}$$

(Note: Here m denotes the Lebesgue measure.)

Solution. For $\lambda > 0$ can divide $\int_E |f| \, \mathrm{d}m$ into

$$\int_E |f| \, \mathrm{d}m = \int_{E \cap \{|f| \leq \lambda\}} |f| \, \mathrm{d}m + \int_{E \cap \{|f| > \lambda\}} |f| \, \mathrm{d}m. \tag{10.194}$$

In the first integral $|f| \leq \lambda$, where the integration is over a set with measure at most $m(E)$. Therefore,

$$\int_{E \cap \{|f| \leq \lambda\}} |f| \, \mathrm{d}m \leq \lambda \, m(E). \tag{10.195}$$

In the second integral of (10.194), we rewrite the integral into

$$\int_{E \cap \{|f| > \lambda\}} |f| \, \mathrm{d}m = \sum_{k=0}^{\infty} \int_{E \cap \left\{\lambda 2^k < |f| \leq \lambda 2^{k+1}\right\}} |f| \, \mathrm{d}m. \tag{10.196}$$

On each summand, $|f| \leq \lambda 2^{k+1}$. Furthermore,

$$m\left(E \cap \left\{\lambda 2^k < |f| \leq \lambda 2^{k+1}\right\}\right) \leq m\left(\left\{\lambda 2^k < |f|\right\}\right) \leq \frac{C}{\lambda^2 2^{2k}}. \tag{10.197}$$

Hence,

$$\int_{E \cap \left\{\lambda 2^k < |f| \leq \lambda 2^{k+1}\right\}} |f| \, \mathrm{d}m \leq \lambda 2^{k+1} \frac{C}{\lambda^2 2^{2k}} = \frac{C}{\lambda} \, 2^{-k+1}. \tag{10.198}$$

Summing up, we get

$$\int_{E \cap \{|f| > \lambda\}} |f| \, \mathrm{d}m \leq \sum_{k=0}^{\infty} \frac{C}{\lambda} \, 2^{-k+1} = \frac{4C}{\lambda}. \tag{10.199}$$

Plugging everything into (10.194), we have

$$\int_E |f| \, \mathrm{d}m \leq \lambda m(E) + \frac{4C}{\lambda}. \tag{10.200}$$

This is true for every $\lambda > 0$. So we minimize the RHS. The RHS is minimized when $\lambda = 2\sqrt{\frac{C}{m(E)}}$, and for this value, the RHS is $4\sqrt{C}\sqrt{m(E)}$. Therefore,

$$\int_E |f| \, \mathrm{d}m \leq 4\sqrt{C} \, \sqrt{m(E)}, \tag{10.201}$$

as desired, where $C_1 = 4\sqrt{C}$. ■

Alternate Solution. One can use The layer-cake formula (to be proven later) to achieve a tighter bound.

$$\int_E |f| \, \mathrm{d}m = \int_0^{\infty} m\left(\{x \in E : |f(x)| > t\}\right) \, \mathrm{d}t \tag{10.202}$$

$$= \int_0^{\lambda} m\left(\{x \in E : |f(x)| > t\}\right) \, \mathrm{d}t + \int_{\lambda}^{\infty} m\left(\{x \in E : |f(x)| > t\}\right) \, \mathrm{d}t. \tag{10.203}$$

The first integral is bounded by

$$\int_0^{\lambda} m\left(\{x \in E : |f(x)| > t\}\right) \, \mathrm{d}t \leq \int_0^{\lambda} m(E) \, \mathrm{d}t = \lambda \, m(E). \tag{10.204}$$

Now, for the second integral

$$\int_\lambda^\infty m\left(\{x \in E : |f(x)| > t\}\right)\ \mathrm{d}t \le \int_\lambda^\infty \frac{C}{\lambda^2}\ \mathrm{d}t = \left[-\frac{C}{t}\right]_\lambda^\infty = \frac{C}{\lambda} \tag{10.205}$$

Therefore, using (10.203), (10.204), (10.205), we have

$$\int_E |f|\ \mathrm{d}m \le \lambda\ m(E) + \frac{C}{\lambda}. \tag{10.206}$$

This is true for every $\lambda > 0$. So we minimize the RHS. The RHS is minimized when $\lambda = \sqrt{\frac{C}{m(E)}}$, and for this value, the RHS is $2\sqrt{C}\sqrt{m\,(E)}$. Therefore,

$$\int_E |f|\ \mathrm{d}m \le 2\sqrt{C}\ \sqrt{m\,(E)}, \tag{10.207}$$

as desired, where $C_1 = 2\sqrt{C}$. ■

Exercise 10.12. Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function in $L^1_{\text{weak}}(\mathbb{R}^d)$, i.e.

$$\theta = \sup_{\lambda>0} \lambda \cdot m(\{x \in \mathbb{R}^d : |f(x)| > \lambda\}) < \infty. \tag{10.208}$$

(here m denotes the Lebesgue measure). Prove that for every $\delta \in (0, 1)$, and for every measurable set $E \subset \mathbb{R}^d$, $m(E) < \infty$, we have

$$\int_E |f(x)|^\delta\ \mathrm{d}x \le C\, m(E)^{1-\delta}\theta^\delta, \tag{10.209}$$

with an appropriate constant C that is independent of f and E.

Solution. We proceed similarly as last problem. For $\lambda > 0$, we write

$$\int_E |f|^\delta\ \mathrm{d}m = \int_{E\cap\{|f|\le\lambda\}} |f|^\delta\ \mathrm{d}m + \sum_{k=0}^{\infty} \int_{E\cap\left\{\lambda 2^k < |f| \le \lambda 2^{k+1}\right\}} |f|^\delta\ \mathrm{d}m. \tag{10.210}$$

In the first integral, the $|f|^\delta \le \lambda^\delta$, and the integration is over a set of measure at most $m(E)$. Therefore,

$$\int_{E\cap\{|f|\le\lambda\}} |f|^\delta\ \mathrm{d}m \le \lambda^\delta \cdot m(E). \tag{10.211}$$

For the second one, we'll use the weak-integrability condition. Given that for every $\lambda > 0$,

$$\lambda \cdot m(\{x \in \mathbb{R}^d : |f(x)| > \lambda\}) \le \theta \implies m(\{x \in \mathbb{R}^d : |f(x)| > \lambda\}) \le \frac{\theta}{\lambda}. \tag{10.212}$$

In the second part of (10.210), on each summand, $|f|^\delta \le \lambda^\delta 2^{(k+1)\delta}$. Furthermore,

$$m\left(E \cap \left\{\lambda 2^k < |f| \le \lambda 2^{k+1}\right\}\right) \le m\left(\left\{\lambda 2^k < |f|\right\}\right) \le \frac{\theta}{\lambda 2^k}. \tag{10.213}$$

Hence,

$$\int_{E\cap\{\lambda 2^k<|f|\le\lambda 2^{k+1}\}} |f|^\delta \, \mathrm{d}m \le \lambda^\delta 2^{(k+1)\delta}\frac{\theta}{\lambda 2^k} = \theta\lambda^{\delta-1}2^\delta 2^{(\delta-1)k} \tag{10.214}$$

Plugging everything into (10.210), we get

$$\int_E |f|^\delta \, \mathrm{d}m \le \lambda^\delta \cdot m(E) + \theta\lambda^{\delta-1}2^\delta \sum_{k=0}^{\infty} 2^{(\delta-1)k} = \lambda^\delta \cdot m(E) + \theta\lambda^{\delta-1}2^\delta \, \frac{1}{1-2^{\delta-1}} \tag{10.215}$$

This is true for any $\lambda > 0$. So we put $\lambda = \frac{\theta}{m(E)}$. Then the RHS becomes

$$\left(\frac{\theta}{m(E)}\right)^\delta m(E) + \theta\left(\frac{\theta}{m(E)}\right)^{\delta-1} \frac{2^\delta}{1-2^{\delta-1}} = \left(1+\frac{2^\delta}{1-2^{\delta-1}}\right)\theta^\delta m(E)^{1-\delta}. \tag{10.216}$$

Hence,

$$\int_E |f|^\delta \, \mathrm{d}m \le \left(1+\frac{2^\delta}{1-2^{\delta-1}}\right)\theta^\delta m(E)^{1-\delta}, \tag{10.217}$$

as desired, with $C = 1 + \frac{2^\delta}{1-2^{\delta-1}}$, which is independent of f and E. ■

Alternate Solution. This can also be proven using The layer-cake formula, and we get a tighter bound.

$$\begin{aligned}\int_E |f(x)|^\delta \, \mathrm{d}m &= \int_0^\infty \delta t^{\delta-1} \, m\left(\{x\in E : |f(x)| > t\}\right) \, \mathrm{d}t \\ &= \int_0^\lambda \delta t^{\delta-1} \, m\left(\{x\in E : |f(x)| > t\}\right) \, \mathrm{d}t \\ &\quad + \int_\lambda^\infty \delta t^{\delta-1} \, m\left(\{x\in E : |f(x)| > t\}\right) \, \mathrm{d}t.\end{aligned} \tag{10.218}$$

The first integral can be bounded by

$$\int_0^\lambda \delta t^{\delta-1} \, m\left(\{x\in E : |f(x)| > t\}\right) \, \mathrm{d}t \le \int_0^\lambda \delta t^{\delta-1} \, m\left(E\right) \, \mathrm{d}t = \lambda^\delta m(E). \tag{10.219}$$

Now, for the second integral,

$$\begin{aligned}\int_\lambda^\infty \delta t^{\delta-1} \, m\left(\{x\in E : |f(x)| > t\}\right) \, \mathrm{d}t &\le \int_\lambda^\infty \delta t^{\delta-1} \, \frac{\theta}{t} \, \mathrm{d}t \\ &= \delta\theta \int_\lambda^\infty t^{\delta-2} \, \mathrm{d}t \\ &= \delta\theta \left[\frac{t^{\delta-1}}{\delta-1}\right]_\lambda^\infty = \frac{\delta}{1-\delta}\theta\lambda^{\delta-1}.\end{aligned} \tag{10.220}$$

Therefore, using (10.218), (10.219), (10.220), we have

$$\int_E |f(x)|^\delta \, \mathrm{d}m \le \lambda^\delta \, m(E) + \frac{\delta}{1-\delta}\theta\lambda^{\delta-1}. \tag{10.221}$$

This is true for every $\lambda > 0$. So we choose $\lambda = \frac{\theta}{m(E)}$. Then the RHS becomes

$$\left(\frac{\theta}{m(E)}\right)^d m(E) + \frac{\delta}{1-\delta}\theta\left(\frac{\theta}{m(E)}\right)^{\delta-1} = \left(1 + \frac{\delta}{1-\delta}\right)\theta^\delta m(E)^{1-\delta}. \tag{10.222}$$

Hence,

$$\int_E |f(x)|^\delta \, \mathrm{d}m \leq \left(1 + \frac{\delta}{1-\delta}\right)\theta^\delta m(E)^{1-\delta}, \tag{10.223}$$

as desired, with $C = 1 + \frac{\delta}{1-\delta} = \frac{1}{1-\delta}$, which is independent of f and E. ■

Exercise 10.13. We say that a sequence $\{f_k\}$ of measurable functions in $\mathbb{R}^d$ is *Cauchy in measure* if for every ε we have

$$m(\{x \in \mathbb{R}^d : |f_k(x) - f_\ell(x)| > \varepsilon\}) \to 0 \quad \text{as } k, \ell \to \infty. \tag{10.224}$$

We say that f_k *converges in measure* to a measurable function f if for every $\varepsilon > 0$ we have

$$m(\{x \in \mathbb{R}^d : |f_k(x) - f(x)| > \varepsilon\}) \to 0 \quad \text{as } k \to \infty. \tag{10.225}$$

Prove the following:

(i) If $\{f_k\}$ is Cauchy in measure, then there is a measurable function f to which this sequence converges in measure.

(ii) If f_k converges to f in measure, show that there exists a subsequence f_{n_k} that converges to f pointwise almost everywhere. Passing to a subsequence is really necessary here?

(iii) If $f_k \to f$ in L^p, then f_k converges to f in measure. Is the converse true? (say, assuming that all f_k and f are in L^p).

Solution. (i) We construct a subsequence $\{g_k\}$ such that

$$m\left(\left\{x : |g_{k+1}(x) - g_k(x)| > 2^{-k}\right\}\right) \leq 2^{-k}. \tag{10.226}$$

Suppose

$$E_k = \left\{x : |g_{k+1}(x) - g_k(x)| > 2^{-k}\right\}, \quad \text{then } \sum_{k=1}^{\infty} m\left(E_k\right) \leq \sum_{k=1}^{\infty} 2^{-k} < \infty. \tag{10.227}$$

By Borel–Cantelli lemma, the set E of points which belong to infinitely many E_k's has measure 0. Now we claim that $\{g_k(x)\}_k$ is a Cauchy sequence for a fixed $x \notin E$.

Since $x \notin E$, there exists K such that $x \notin E_k$ for $k \geq K$. Now, for $n > m \geq K + 1$,

$$|g_n(x) - g_m(x)| \leq \sum_{i=m}^{n-1} |g_{i+1}(x) - g_i(x)| \leq \sum_{i=m}^{n-1} 2^{-i} \leq \sum_{i=m}^{\infty} 2^{-i} = 2^{-m+1} \leq 2^{-K}. \tag{10.228}$$

Therefore, $\{g_k(x)\}_k$ is a Cauchy sequence. Hence, by the completeness of $\mathbb{R}$ (or $\mathbb{C}$), we can define

$$f(x) = \lim_{n\to\infty} g_n(x), \qquad \text{for } x \notin E. \tag{10.229}$$

For $x \in E$, we can just define $f(x) = 0$. Now we'll prove that $g_k \to f$ in measure. Indeed, for $x \notin \bigcup_{n=k}^{\infty} E_n$,

$$|f(x) - g_k(x)| = \lim_{n\to\infty} |g_n(x) - g_k(x)| \le 2^{-k+1}. \tag{10.230}$$

Therefore, if $|f(x) - g_k(x)| > 2^{-k+1}$, then x must belong to $\bigcup_{n=k}^{\infty} E_n$, or the nullset E. Hence,

$$\begin{aligned} m\left(\left\{x : |f(x) - g_k(x)| > 2^{-k+1}\right\}\right) &\le m\left(A \cup \bigcup_{n=k}^{\infty} E_n\right) \\ &\le \sum_{n=k}^{\infty} m(E_n) \\ &\le \sum_{n=k}^{\infty} 2^{-n} \le 2^{-k+1}. \end{aligned} \tag{10.231}$$

Hence, $g_n \to f$ in measure. Now we claim that $f_n \to f$ in measure. Notice that

$$|f(x) - f_n(x)| \le |f(x) - g_n(x)| + |g_n(x) - f_n(x)|. \tag{10.232}$$

This gives

$$\{x : |f(x) - f_n(x)| \ge \varepsilon\} \subseteq \left\{x : |f(x) - g_n(x)| \ge \frac{\varepsilon}{2}\right\} \cup \left\{x : |g_n(x) - f_n(x)| \ge \frac{\varepsilon}{2}\right\}. \tag{10.233}$$

As $n \to \infty$, the first set's measure approaches 0, since $g_n \to f$ in measure. As $n \to \infty$, the second set's measure also approaches 0, since f_n is a Cauchy sequence and g_n is a subsequence of f_n. Therefore,

$$\lim_{n\to\infty} m(\{x : |f(x) - f_n(x)| \ge \varepsilon\}) = 0, \tag{10.234}$$

i.e. $f_n \to f$ in measure.

(ii) We proceed similarly as before. We construct a subsequence $\{g_k\}$ as follows: for a fixed k, there exists N_k such that for all $m \ge N_k$,

$$m\left(\left\{x : |f_m(x) - f(x)| > 2^{-k}\right\}\right) \le 2^{-k}. \tag{10.235}$$

Then $g_k = f_{N_k}$. Suppose

$$E_k = \left\{x : |g_k(x) - f(x)| > 2^{-k}\right\}, \qquad \text{then } \sum_{k=1}^{\infty} m(E_k) \le \sum_{k=1}^{\infty} 2^{-k} < \infty. \tag{10.236}$$

By Borel–Cantelli lemma, the set E of points which belong to infinitely many E_k's has measure 0. Now we claim that $g_k \to f$ pointwise on $\mathbb{R}^d - E$.

For $x \notin E$, there exists K such that $x \notin E_k$ for $k \geq K$. Then for all $k \geq K$,

$$|g_k(x) - f(x)| \leq 2^{-k}. \tag{10.237}$$

Therefore, $g_k \to f$ pointwise a.e.

Yes, passing to a subsequence is necessary. For example, consider the sequence of functions $[0,1] \to \mathbb{R}$ inspired from Exercise 6.1:

$$\begin{gathered} f_1 = \chi_{[0,1]} \\ f_2 = \chi_{[0,\frac{1}{2}]},\ f_3 = \chi_{[\frac{1}{2},1]} \\ f_4 = \chi_{[0,\frac{1}{4}]},\ f_5 = \chi_{[\frac{1}{4},\frac{1}{2}]},\ f_6 = \chi_{[\frac{1}{2},\frac{3}{4}]},\ f_7 = \chi_{[\frac{3}{4},1]} \\ \dots \end{gathered}$$

In other words, for $n = 2^k + m$, with k being the unique number satisfying $2^k \leq n < 2^{k+1}$, and $0 \leq m < 2^k$, we define

$$f_n = \chi_{\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right]}. \tag{10.238}$$

Then the pointwise sequence $f_n(x)$ doesn't converge for any x, since the sequence is a sequence of 0 and 1, which is not eventually constant, and hence it doesn't converge. However, it converges to 0 in measure. Given $\varepsilon > 0$

$$m\left(x : |f_n(x) - 0| > \varepsilon\right) = m\left(\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right]\right) = \frac{1}{2^k} \to 0. \tag{10.239}$$

Therefore, a sequence of functions converging in measure may fail to converge pointwise.

(iii) Fix $\varepsilon > 0$. We need to show that

$$\lim_{n\to\infty} m\left(\{x : |f_n(x) - f(x)| > \varepsilon\}\right) = 0. \tag{10.240}$$

Now,

$$\begin{aligned} \|f_n - f\|_p^p &= \int_{\mathbb{R}^d} |f_n - f|^p \, \mathrm{d}m \\ &\geq \int_{\{x : |f_n(x) - f(x)| > \varepsilon\}} |f_n - f|^p \, \mathrm{d}m \\ &\geq \int_{\{x : |f_n(x) - f(x)| > \varepsilon\}} \varepsilon^p \, \mathrm{d}m \\ &= \varepsilon^p m\left(\{x : |f_n(x) - f(x)| > \varepsilon\}\right). \end{aligned}$$

In other words,

$$m\left(\{x : |f_n(x) - f(x)| > \varepsilon\}\right) \leq \frac{\|f_n - f\|_p^p}{\varepsilon^p}. \tag{10.241}$$

$\|f_n - f\|_p^p \to 0$ as $n \to \infty$. Therefore,

$$\lim_{n\to\infty} m\left(\{x : |f_n(x) - f(x)| > \varepsilon\}\right) = 0, \tag{10.242}$$

i.e. $f_n \to f$ in measure.

No, the converse is not true, in general. Suppose

$$f_n = n^{\frac{1}{p}} \chi_{\left(0,\frac{1}{n}\right)}. \tag{10.243}$$

Then $f_n \not\to 0$ in L^p, since

$$\|f_n - 0\|^p = \int_{\mathbb{R}} \left[n^{\frac{1}{p}} \chi_{\left(0,\frac{1}{n}\right)}\right]^p \, \mathrm{d}m = n \cdot m\left(\left(0, \frac{1}{n}\right)\right) = 1. \tag{10.244}$$

However, $f_n \to 0$ in measure. Because,

$$m\left(x : |f_n(x) - 0| > \varepsilon\right) = m\left(\left(0, \frac{1}{n}\right)\right) = \frac{1}{n} \to 0. \tag{10.245}$$

■

Exercise 10.14. Assume that (Ω, Σ, μ) is a measure space and $f \in L^1(\Omega, \Sigma, \mu)$ satisfies $\|f\|_1 = 1$. Prove that

$$\int_E \log|f(x)| \, \mathrm{d}\mu(x) \leq -\mu(E) \log \mu(E) \tag{10.246}$$

for any measurable subset E such that $0 < \mu(E) < \infty$.

Solution. If $f = 0$ in a positive measure set, then $\log|f(x)| = -\infty$ on a positive measure set, implying that the LHS is $-\infty$, in which case the inequality is satisfied trivially. So we can assume $|f| > 0$ a.e. We shall use the bound

$$\log x \leq x - 1, \quad \text{for } x > 0. \tag{10.247}$$

As a result,

$$\log\left(|f(x)| \, \mu(E)\right) \leq |f(x)| \, \mu(E) - 1 \implies \log|f(x)| \leq |f(x)| \, \mu(E) - 1 - \log\left(\mu(E)\right). \tag{10.248}$$

Integrating both sides, we get

$$\int_E \log|f| \, \mathrm{d}\mu \leq \mu(E) \int_E |f| \, \mathrm{d}\mu - \int_E \mathrm{d}\mu - \log\left(\mu(E)\right) \int_E \mathrm{d}\mu \tag{10.249}$$

$$\leq \mu(E) \int_E |f| \, \mathrm{d}\mu - \mu(E) - \mu(E) \log\left(\mu(E)\right). \tag{10.250}$$

Now,

$$\int_E |f| \, \mathrm{d}\mu \leq \int_\Omega |f| \, \mathrm{d}\mu = 1, \tag{10.251}$$

so we have

$$\int_E \log|f| \, \mathrm{d}\mu \leq \mu(E) - \mu(E) - \mu(E) \log\left(\mu(E)\right) = -\mu(E) \log\left(\mu(E)\right). \tag{10.252}$$

■

Alternate Solution. This can also be solved using Jensen's inequality for concave functions (which will be proven soon). log is a concave function, so we have

$$\log\left[\int_E |f|\,\mathrm{d}\nu\right] \geq \int_E \log|f|\,\mathrm{d}\nu, \tag{10.253}$$

where ν is a probability measure on E, i.e. $\nu(E) = 1$. We can just take

$$\nu(A) = \frac{\mu(A)}{\mu(E)}, \qquad \text{for measurable } A \subseteq E. \tag{10.254}$$

Claim 10.2: $\int_E g\,\mathrm{d}\nu = \frac{1}{\mu(E)}\int_E g\,\mathrm{d}\mu$ for measurable functions g.

Proof. It suffices to prove it for $g \in \mathcal{M}^+$, because otherwise we can just apply it on g^+ and g^- and then add up. First, for simple functions $\varphi = \sum_{i=1}^n c_i\chi_{A_i}$, we have

$$\int_E \varphi\,\mathrm{d}\nu = \sum_{i=1}^n c_i\nu(A_i) = \frac{1}{\mu(E)}\sum_{i=1}^n c_i\mu(A_i) = \frac{1}{\mu(E)}\int_E \varphi\,\mathrm{d}\mu. \tag{10.255}$$

Now, for a general $g \in \mathcal{M}^+$,

$$\int_E g\,\mathrm{d}\nu = \sup_{\text{simple }\varphi\leq g}\int_E \varphi\,\mathrm{d}\nu = \sup_{\text{simple }\varphi\leq g}\frac{1}{\mu(E)}\int_E \varphi\,\mathrm{d}\mu = \frac{1}{\mu(E)}\int_E g\,\mathrm{d}\mu. \tag{10.256}$$

□

Now, the RHS of (10.253) is

$$\int_E \log|f|\,\mathrm{d}\nu = \frac{1}{\mu(E)}\int_E \log|f|\,\mathrm{d}\mu. \tag{10.257}$$

Now, the LHS of (10.253) is

$$\begin{aligned}
\log\left[\int_E |f|\,\mathrm{d}\nu\right] &= \log\left[\frac{1}{\mu(E)}\int_E |f|\,\mathrm{d}\mu\right] && (10.258)\\
&= \log\left[\int_E |f|\,\mathrm{d}\mu\right] - \log(\mu(E)) && (10.259)\\
&\leq \log\left[\int_\Omega |f|\,\mathrm{d}\mu\right] - \log(\mu(E)) && (10.260)\\
&= \log 1 - \log(\mu(E)) = -\log(\mu(E)). && (10.261)
\end{aligned}$$

Now, combining (10.257) and (10.261) with (10.253), we have

$$-\log(\mu(E)) \geq \frac{1}{\mu(E)}\int_E \log|f|\,\mathrm{d}\mu \implies \int_E \log|f|\,\mathrm{d}\mu \leq -\mu(E)\log(\mu(E)). \tag{10.262}$$

■

Exercise 10.15. Let $\{f_n\}_{n=1}^{\infty} \subset L^2([0,1])$. Assume that $\sup_n \|f_n\|_2 < \infty$ and let $1 \leq p < 2$. Let $f : [0,1] \to \mathbb{R}$ and assume that $f_n \to f$ pointwise a.e. Prove that

$$\lim_{n\to\infty} \|f_n - f\|_p = 0. \tag{10.263}$$

Give an example to show that this is false for $p = 2$.

Solution. Suppose $M := \sup_n \|f_n\|_2$. By Fatou's Lemma, since $|f_n|^2 \to |f|^2$ pointwise a.e.

$$\int_{[0,1]} |f|^2 \, \mathrm{d}m \leq \liminf_{n\to\infty} \int_{[0,1]} |f_n|^2 \, \mathrm{d}m \leq \liminf_{n\to\infty} M^2 \leq M^2, \tag{10.264}$$

proving that $\|f\|_2 \leq M$. Therefore,

$$\|f_n - f\|_2 \leq \|f_n\|_2 + \|f\|_2 \leq 2M. \tag{10.265}$$

Now, fix $\varepsilon > 0$. By Egorov's theorem, there exists a measurable subset $E \subseteq [0,1]$ such that $f_n \to f$ uniformly on $[0,1] - E$, with

$$m(E) \leq \delta := \left(\frac{\varepsilon^2}{4(2M)^{2p}}\right)^{\frac{1}{2-p}}. \tag{10.266}$$

Since $f_n \to f$ uniformly on $[0,1] - E$, there exists N such that for $n \geq N$,

$$\sup_{x\in[0,1]-E} |f_n(x) - f(x)| \leq \left(\frac{\varepsilon}{2}\right)^{\frac{1}{p}}. \tag{10.267}$$

Now, for $n \geq N$,

$$\|f_n - f\|_p^p = \int_{[0,1]} |f_n - f|^p \, \mathrm{d}m = \int_{[0,1]-E} |f_n - f|^p \, \mathrm{d}m + \int_E |f_n - f|^p \, \mathrm{d}m. \tag{10.268}$$

On $[0,1] - E$, the integrand $|f_n - f|^p$ is at most $\frac{\varepsilon}{2}$, and the integration is over a set of measure at most 1. Therefore

$$\int_{[0,1]-E} |f_n - f|^p \, \mathrm{d}m \leq \frac{\varepsilon}{2}. \tag{10.269}$$

By Hölder's inequality on $\frac{2}{p}$ and its conjugate exponent $s = \frac{2}{2-p}$,

$$\begin{aligned}\int_E |f_n - f|^p \, \mathrm{d}m &\leq \left(\int_E \left(|f_n - f|^p\right)^{\frac{2}{p}} \, \mathrm{d}m\right)^{\frac{p}{2}} \left(\int_E 1^{\frac{2}{2-p}} \, \mathrm{d}m\right)^{\frac{2-p}{2}} \\ &\leq \left(\int_{[0,1]} |f_n - f|^2 \, \mathrm{d}m\right)^{\frac{p}{2}} [m(E)]^{\frac{2-p}{2}} \\ &\leq \left((2M)^2\right)^{\frac{p}{2}} \delta^{\frac{2-p}{2}} \\ &\leq (2M)^p \left(\frac{\varepsilon^2}{4(2M)^{2p}}\right)^{\frac{1}{2}} = \frac{\varepsilon}{2}. \end{aligned} \tag{10.270}$$

Plugging everything into (10.268), we have

$$\|f_n - f\|_p^p \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \quad \text{for } n \geq N. \tag{10.271}$$

Hence, $\lim_{n\to\infty} \|f_n - f\|_p^p = 0$.

For $p = 2$, consider the following example:

$$f_n = \sqrt{n}\chi_{\left(0,\frac{1}{n}\right)}. \tag{10.272}$$

Then

$$\int_{[0,1]} |f_n|^2 \, \mathrm{d}m = n \cdot m\left(\left(0, \frac{1}{n}\right)\right) = 1, \tag{10.273}$$

so that $f_n \in L^2([0,1])$ with $\|f_n\|_2 = 1$, in particular, $\sup_n \|f_n\|_2 = 1 < \infty$. Now, $f_n \to 0$ pointwise, but $\|f_n - 0\|_2 = 1$, so $\lim_{n\to\infty} \|f_n - 0\|_2 \neq 0$. ■

Exercise 10.16. Let $\psi : \mathbb{R} \to \mathbb{R}$ be a bounded and continuous function such that

$$\left|\int_a^b \psi(x) \, \mathrm{d}x\right| \leq C \quad \text{for any } a, b \in \mathbb{R}. \tag{10.274}$$

If $f \in L^1(\mathbb{R})$, consider the function $\widetilde{f}$ defined on $\mathbb{R}$ by

$$\widetilde{f}(y) := \int_{\mathbb{R}} f(x)\psi(yx) \, \mathrm{d}x. \tag{10.275}$$

(i) Show that $\widetilde{f}$ is a continuous function.

(ii) Show that

$$\lim_{|y|\to\infty} \widetilde{f}(y) = 0. \tag{10.276}$$

Solution. (i) Let y_n be a sequence converging to y. Then $\psi(y_n x)$ converges to $\psi(yx)$ for every x, since ψ is continuous. Also note that since ψ is bounded, $|\psi(z)| \leq M$ for every z. Therefore,

$$|f(x)\psi(y_n x)| \leq |f(x)| M, \tag{10.277}$$

and $M|f| \in L^1$. Therefore, by Dominated Convergence Theorem,

$$\lim_{n\to\infty} \widetilde{f}(y_n) = \lim_{n\to\infty} \int_{\mathbb{R}} f(x)\psi(y_n x)\,\mathrm{d}x = \int_{\mathbb{R}} f(x) \lim_{n\to\infty} \psi(y_n x)\,\mathrm{d}x \tag{10.278}$$

$$= \int_{\mathbb{R}} f(x)\psi(yx)\,\mathrm{d}x = \widetilde{f}(y). \tag{10.279}$$

Therefore, $\widetilde{f}$ is continuous.

(ii) Note that

$$\widetilde{cf} = c\widetilde{f}, \text{ and } \quad \widetilde{f_1 + f_2} = \widetilde{f}_1 + \widetilde{f}_2. \tag{10.280}$$

It's easy to see this:

$$\widetilde{cf}(y) = \int_{\mathbb{R}} (cf)(x)\psi(yx)\,\mathrm{d}x = c\int_{\mathbb{R}} f(x)\psi(yx)\,\mathrm{d}x = c\widetilde{f}(y), \tag{10.281}$$

$$\begin{aligned}\widetilde{f_1 + f_2}(y) &= \int_{\mathbb{R}} (f_1 + f_2)(x)\psi(yx)\,\mathrm{d}x \\ &= \int_{\mathbb{R}} f_1(x)\psi(yx)\,\mathrm{d}x + \int_{\mathbb{R}} f_2(x)\psi(yx)\,\mathrm{d}x \\ &= \widetilde{f_1}(y) + \widetilde{f_2}(y).\end{aligned} \tag{10.282}$$

Therefore, if (10.276) holds for $f \in L^1$, it also holds for $cf \in L^1$. Furthermore, if it holds for $f_1, f_2 \in L^1$, it also holds for $f_1 + f_2$.

Now, we'll show first it holds for $f = \chi_{[a,b]}$. Then

$$\widetilde{f}(y) = \int_{\mathbb{R}} \chi_{[a,b]}(x)\psi(yx)\,\mathrm{d}x = \int_a^b \chi_{[a,b]}(x)\psi(yx)\,\mathrm{d}x = \frac{1}{|y|}\int_{a'}^{b'} \psi(u)\,\mathrm{d}u, \tag{10.283}$$

for $u = yx$. Then we have

$$\left|\widetilde{f}(y)\right| = \frac{1}{|y|}\left|\int_{a'}^{b'} \psi(u)\,\mathrm{d}u\right| \le \frac{C}{|y|} \to 0 \text{ as } |y| \to \infty. \tag{10.284}$$

Therefore, (10.276) holds for $f = \chi_{[a,b]}$. By linearity, it holds for any step function. Now, let $f \in L^1$ be arbitrary. Then we can find a step function g such that

$$\|f - g\| \le \frac{\varepsilon}{2M}. \tag{10.285}$$

Then we have

$$\begin{aligned}\left|\widetilde{f}(y) - \widetilde{g}(y)\right| &= \left|\int_{\mathbb{R}} (f(x) - g(x))\,\psi(yx)\right| \\ &\le M\int_{\mathbb{R}} |f(x) - g(x)|\,\mathrm{d}x \\ &= M\|f - g\|_1 \le \frac{\varepsilon}{2}.\end{aligned} \tag{10.286}$$

Now, since $\lim_{|y|\to\infty} \widetilde{g}(y) = 0$, we can find N such that $|y| > N$ implies $|\widetilde{g}(y)| \le \frac{\varepsilon}{2}$. Then

we have for $|y| > N$,

$$\left|\widetilde{f}(y)\right| \leq \left|\widetilde{f}(y) - \widetilde{g}(y)\right| + |\widetilde{g}(y)| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \tag{10.287}$$

proving that $\lim_{|y|\to\infty} \widetilde{f}(y) = 0$.

■

Exercise 10.17. Prove or show a counterexample: if $f \in L^1(\mathbb{R}) \cap L^3(\mathbb{R})$ then $f \in L^2(\mathbb{R})$.

Solution. Yes this is true. In fact this is true for every measure space (Ω, Σ, μ).

$$\int_\Omega |f|^2 \, \mathrm{d}\mu = \int_{\{|f|\leq 1\}} |f|^2 \ \mathrm{d}\mu + \int_{\{|f|>1\}} |f|^2 \ \mathrm{d}\mu. \tag{10.288}$$

If $|f| \leq 1$, then $|f|^2 \leq |f|^1$; and if $|f| > 1$, then $|f|^2 \leq |f|^3$. Therefore,

$$\int_\Omega |f|^2 \, \mathrm{d}\mu \leq \int_{\{|f|\leq 1\}} |f| \ \mathrm{d}\mu + \int_{\{|f|>1\}} |f|^3 \ \mathrm{d}\mu \leq \int_\Omega |f| \, \mathrm{d}\mu + \int_\Omega |f|^3 \, \mathrm{d}\mu < \infty. \tag{10.289}$$

Hence, $f \in L^2$. ■

§11. 28 September 2026

Throughout, (Ω, Σ, μ) is a measure space. For $1 \le p \le \infty$, the *conjugate exponent* p' is defined by

$$\frac{1}{p} + \frac{1}{p'} = 1, \qquad \text{i.e.} \qquad p' = \frac{p}{p-1}. \tag{11.1}$$

Recall that L^p is a Banach space. We assume that μ is σ-finite, that is,

$$\Omega = \bigcup_{n=1}^{\infty} \Omega_n, \qquad \mu(\Omega_n) < \infty \text{ for all } n. \tag{11.2}$$

For example, Lebesgue measure on $\mathbb{R}^d$ is σ-finite. This hypothesis is needed in the case $p = \infty$ below.

§11.i. Duality of the norm

Proposition 11.1 (Duality of the norm)

Let $f \in L^p(\Omega)$ with $1 \le p \le \infty$. Then

$$\|f\|_p = \sup \left\{ \left| \int_\Omega fg \, \mathrm{d}\mu \right| : g \in L^{p'}(\Omega),\ \|g\|_{p'} \le 1 \right\}. \tag{11.3}$$

(For real-valued functions the absolute value may be dropped.) Moreover, if $1 \le p < \infty$, the supremum is always attained, so it is a maximum.

This is the statement, in the language of norms, that $L^{p'}$ is the dual space of L^p.

Proof. By Hölder's inequality, if $g \in L^{p'}$ with $\|g\|_{p'} \le 1$, then

$$\left| \int fg \, d\mu \right| \le \|f\|_p \, \|g\|_{p'} \le \|f\|_p . \tag{11.4}$$

Hence the supremum is at most $\|f\|_p$.

Now suppose $1 \le p < \infty$. We look for a good g, one that "resonates" with the sign of f and makes fg comparable to $|f|^p$. Since fg should carry one power of f and we want $|f|^p$, we take $|f|^{p-1}$ times the sign of f. Define

$$g(x) = \frac{\operatorname{sgn}(f(x)) \, |f(x)|^{p-1}}{\|f\|_p^{p-1}}, \tag{11.5}$$

where

$$\operatorname{sgn}(t) = \begin{cases} 1 & \text{if } t > 0, \\ -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0. \end{cases} \tag{11.6}$$

For complex case, we just define

$$\operatorname{sgn}(z) = \begin{cases} \frac{|z|}{z} & \text{if } z \neq 0, \\ 0 & \text{if } z = 0. \end{cases} \tag{11.7}$$

Claim 11.1: $g \in L^{p'}$ and $\|g\|_{p'} = 1$.

Proof. Since $p' = \frac{p}{p-1}$, we have $(p-1)p' = p$, so

$$\int_\Omega |g|^{p'} \, \mathrm{d}\mu = \frac{1}{\|f\|_p^{(p-1)p'}} \int_\Omega |f|^{(p-1)p'} \, \mathrm{d}\mu = \frac{1}{\|f\|_p^p} \int_\Omega |f|^p \, \mathrm{d}\mu = 1. \tag{11.8}$$

In particular the integral is finite because $f \in L^p$. □

Claim 11.2: g attains the supremum.

Proof. Since $f \operatorname{sgn}(f) = |f|$,

$$\int_\Omega fg \, \mathrm{d}\mu = \frac{1}{\|f\|_p^{p-1}} \int_\Omega |f| \, |f|^{p-1} \, \mathrm{d}\mu = \frac{\|f\|_p^p}{\|f\|_p^{p-1}} = \|f\|_p . \tag{11.9}$$

□

This proves the proposition for $1 \leq p < \infty$, and shows the supremum is attained.

The case $p = \infty$ (so $p' = 1$). Let $\varepsilon > 0$ and set

$$X_\varepsilon = \{x \in \Omega : |f(x)| \geq \|f\|_\infty - \varepsilon\}. \tag{11.10}$$

By the definition of the essential supremum, $\mu(X_\varepsilon) > 0$ (otherwise $\|f\|_\infty - \varepsilon$ would be a smaller essential upper bound). However, $\mu(X_\varepsilon)$ could be $+\infty$. This is where σ-finiteness is used: writing

$$X_\varepsilon = \bigcup_n (X_\varepsilon \cap \Omega_n), \tag{11.11}$$

at least one piece has positive measure, otherwise $\mu(X_\varepsilon) = 0$. So there is a set $Y_\varepsilon \subseteq X_\varepsilon$ with $0 < \mu(Y_\varepsilon) < \infty$. Define

$$g = \frac{\operatorname{sgn}(f) \, \chi_{Y_\varepsilon}}{\mu(Y_\varepsilon)}. \tag{11.12}$$

Then $\|g\|_1 = 1$, since

$$\int_\Omega |g| \, \mathrm{d}\mu = \frac{1}{\mu(Y_\varepsilon)} \int_\Omega |\operatorname{sgn}(f)| \, \chi_{Y_\varepsilon} \, \mathrm{d}\mu = \frac{1}{\mu(Y_\varepsilon)} \int_\Omega \chi_{Y_\varepsilon} \, \mathrm{d}\mu = 1. \tag{11.13}$$

Then

$$\int_\Omega fg \, \mathrm{d}\mu = \frac{1}{\mu(Y_\varepsilon)} \int_{Y_\varepsilon} |f| \, \mathrm{d}\mu \geq \|f\|_\infty - \varepsilon. \tag{11.14}$$

Since $\varepsilon > 0$ was arbitrary, the supremum equals $\|f\|_\infty$. ■

Remark 11.1. Note that, the whole proof works if we replace $\|g\|_{p'} \leq 1$ with $\|g\|_{p'} = 1$. Hence,

$$\|f\|_p = \sup_{\substack{g \in L^{p'}(\Omega), \\ \|g\|_{p'}=1}} \left| \int_\Omega fg \, \mathrm{d}\mu \right|. \tag{11.15}$$

We can further restrict our attention to simple functions for $1 < p \leq \infty$, since for that case, simple functions are dense in $L^{p'}(\Omega)$. So we can find a sequence of simple functions g_n and

$$g_n \to g \text{ in } L^{p'}. \tag{11.16}$$

Then by Hölder's inequality,

$$\left| \int_\Omega fg \, \mathrm{d}\mu - \int_\Omega fg_n \, \mathrm{d}\mu \right| \leq \int_\Omega |f(g - g_n)| \, \mathrm{d}\mu \leq \|f\|_p \|g - g_n\|_{p'} \to 0 \tag{11.17}$$

as $n \to \infty$. So we have that

$$\int_\Omega fg \, \mathrm{d}\mu = \lim_{n\to\infty} \int_\Omega fg_n \, \mathrm{d}\mu, \tag{11.18}$$

i.e. any $\int_\Omega fg \, \mathrm{d}\mu$ can be approximated by $\int_\Omega fg_n \, \mathrm{d}\mu$ where g_n's are simple functions. As a result,

$$\sup_{\substack{g \in L^{p'}(\Omega), \\ \|g\|_{p'}=1}} \left| \int_\Omega fg \, \mathrm{d}\mu \right| = \sup_{\substack{g \text{ simple}, \\ \|g\|_{p'}=1}} \left| \int_\Omega fg \, \mathrm{d}\mu \right|. \tag{11.19}$$

Remark 11.2. The supremum need not be attained for $p = \infty$. If $|f|$ has a plateau at height $\|f\|_\infty$ (on a set of positive finite measure), one can take g supported on that plateau and the supremum is attained. The opposite situation is a function that reaches its maximum at a single point, for instance $f(x) = (1 - |x|)^+$ on $\mathbb{R}$ with $\|f\|_\infty = 1$.

To have the supermum attained, we need to have $\int_\mathbb{R} fg \, \mathrm{d}m = 1$, with $\|g\|_1 \leq 1$. Now, notice that since $f < 1$ a.e., we have

$$1 = \left| \int_\mathbb{R} fg \, \mathrm{d}m \right| \leq \int_\mathbb{R} |f| \, |g| \, \mathrm{d}m < \int_\mathbb{R} |g| \, \mathrm{d}m = \|g\|_1 \leq 1, \tag{11.20}$$

which is a contradiction! Therefore, the supremum need not be attained for $p = \infty$.

Remark 11.3 (Compactness of the unit ball; weak topologies)**.** The unit ball in an infinite-dimensional normed space is not compact in the norm topology. For example, on $\Omega = \mathbb{N}$ with counting measure, $L^p = \ell^p$ is the space of sequences (a_n) with $\sum |a_n|^p < \infty$. Let e_k be the sequence with 1 in position k and 0 elsewhere. Then $\|e_k\|_p = 1$ and for $j \neq k$,

$$\|e_j - e_k\|_p = 2^{1/p} > 0, \tag{11.21}$$

so (e_k) has no convergent subsequence. Since ℓ^p is a metric space, compactness is equivalent to sequential compactness, so the closed unit ball is not compact. (Compact means every open cover has a finite subcover; "closed and bounded" is a property of compact sets in $\mathbb{R}^d$, not the definition.)

This motivates *weak topologies*: a topology with fewer open sets than the norm (strong) topology, hence more compact sets, but still meaningful. This is useful, for example, for producing solutions of PDEs by minimizing a functional: one needs a minimizing sequence to have a convergent subsequence. This will be discussed in a functional analysis course.

§11.ii. Distribution function and weak L^p

Definition 11.1. Let $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) be measurable. For $\lambda > 0$ define the **super-level sets**

$$E_\lambda = \{x \in \Omega : |f(x)| > \lambda\}. \tag{11.22}$$

If $\lambda_1 > \lambda_2$, then $E_{\lambda_1} \subseteq E_{\lambda_2}$. The function $\lambda \mapsto \mu(E_\lambda)$ is called the **distribution function** of f.

The distribution function encodes much of the information about f. For instance, let $f \in L^p$ with $1 \le p < \infty$. On E_λ we have $|f| > \lambda$, so

$$\int_\Omega |f|^p \, \mathrm{d}\mu \ge \int_{E_\lambda} |f|^p \, \mathrm{d}\mu \ \ge \ \lambda^p \, \mu(E_\lambda). \tag{11.23}$$

Hence,

$$\mu(E_\lambda) \ \le \ \frac{\|f\|_p^p}{\lambda^p} \qquad \text{for all } \lambda > 0. \tag{11.24}$$

So the super-level sets cannot be large when $\|f\|_p$ is fixed. This is called Chebyshev's inequality. (This is the generalization of Markov's inequality in probability language.)

Proposition 11.2 (Chebyshev's inequality)

Let $f \in L^p$ with $1 \le p < \infty$. Then

$$\mu\left(\{x \in \Omega : |f(x)| > \lambda\}\right) \ \le \ \frac{\|f\|_p^p}{\lambda^p} \qquad \text{for all } \lambda > 0. \tag{11.25}$$

This motivates the definition of weak L^p spaces.

Definition 11.2 (Weak L^p)**.** Let $1 \le p < \infty$. The space L^p_{weak} (also written $L^{p,\infty}$) is the set of measurable functions $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) for which there exists a constant $C > 0$, depending on f (and p) but *not* on λ, such that

$$\mu\left(\{x \in \Omega : |f(x)| > \lambda\}\right) \le \frac{C}{\lambda^p} \qquad \text{for all } \lambda > 0. \tag{11.26}$$

Remark 11.4. By (11.24), every $f \in L^p$ satisfies this with $C = \|f\|_p^p$, so

$$L^p \subset L^p_{\text{weak}}. \tag{11.27}$$

The inclusion is strict. Let $f(x) = |x|^{-d}$ on $\mathbb{R}^d$. Then

$$E_\lambda = \{x : |x|^{-d} > \lambda\} = \left\{x : |x| < \lambda^{-\frac{1}{d}}\right\}, \tag{11.28}$$

a ball of radius $\lambda^{-1/d}$, so

$$\mu(E_\lambda) = \omega_d \left(\lambda^{-1/d}\right)^d = \frac{\omega_d}{\lambda}, \tag{11.29}$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$. Hence $f \in L^1_{\text{weak}}(\mathbb{R}^d)$.

But $f \notin L^1(\mathbb{R}^d)$. In polar coordinates,

$$\int_{|x|<1} \frac{dx}{|x|^d} \sim \int_0^1 \frac{1}{r^d}\, r^{d-1}\, dr = \int_0^1 \frac{dr}{r} = \infty, \qquad \int_{|x|>1} \frac{dx}{|x|^d} \sim \int_1^\infty \frac{dr}{r} = \infty. \tag{11.30}$$

So $|x|^{-d}$ is exactly the threshold function for integrability: it fails to be integrable both near 0 and near ∞, but only barely.

§11.iii. Integrability of powers, and how to test $f \in L^p$

For $\alpha > 0$, on $\mathbb{R}^d$,

$$\int_{|x|<1} \frac{dx}{|x|^\alpha} < \infty \iff \alpha < d, \qquad \int_{|x|>1} \frac{dx}{|x|^\alpha} < \infty \iff \alpha > d. \tag{11.31}$$

For a function that is continuous away from the origin, to decide whether $f \in L^p$ one only needs its asymptotic behavior near 0 and near ∞.

Example 11.1

On $\mathbb{R}^3$ let $f(x) = \dfrac{1}{|x|\,(1+|x|)}$. Near 0, $f \sim |x|^{-1}$, so $f^p \sim |x|^{-p}$ is integrable near 0 iff $p < 3$. Near ∞, $f \sim |x|^{-2}$, so $f^p \sim |x|^{-2p}$ is integrable near ∞ iff $2p > 3$. Hence

$$f \in L^p(\mathbb{R}^3) \iff \frac{3}{2} < p < 3. \tag{11.32}$$

Remark 11.5. Power functions are not the only threshold. On $[2, \infty)$,

$$\int \frac{\mathrm{d}x}{x \log x} = \log\log x \tag{11.33}$$

diverges, while

$$\int \frac{\mathrm{d}x}{x(\log x)^{1+\varepsilon}} < \infty \tag{11.34}$$

§11.iv. The layer-cake formula

Proposition 11.3 (The layer-cake formula)

Let $f \in L^p$, $1 \le p < \infty$.

$$\|f\|_p^p = \int_0^\infty pt^{p-1}\,\mu(E_t)\,\mathrm{d}t. \tag{11.35}$$

Proof. Let $f \in L^p$, $1 \le p < \infty$. Since

$$|f(x)|^p = \int_0^{|f(x)|} pt^{p-1}\,\mathrm{d}t, \tag{11.36}$$

we can write

$$\|f\|_p^p = \int_\Omega |f|^p\,\mathrm{d}\mu = \int_\Omega \int_0^\infty pt^{p-1}\,\chi_{\{t<|f(x)|\}}\,\mathrm{d}t\,\mathrm{d}\mu. \tag{11.37}$$

The integrand is nonnegative and measurable, so by Tonelli's theorem (interchanging integrals of nonnegative functions is always allowed, though the value may be $+\infty$, we'll prove this result soon),

$$\|f\|_p^p = \int_\Omega |f|^p\,\mathrm{d}\mu = \int_0^\infty pt^{p-1} \int_\Omega \chi_{\{t<|f(x)|\}}\,\mathrm{d}\mu\,\mathrm{d}t = \int_0^\infty pt^{p-1}\,\mu(E_t)\,\mathrm{d}t. \tag{11.38}$$

■

Consequently, the norm of f is determined by its distribution function, and two functions with the same distribution function have the same L^p norm. Similarly, for a measurable set $E \subset \Omega$,

$$\int_E |f|^p\,\mathrm{d}\mu = \int_0^\infty pt^{p-1}\,\mu(E_t \cap E)\,\mathrm{d}t. \tag{11.39}$$

§11.v. Jensen's inequality

Definition 11.3. Let (a, b) be an interval (one may have $a = -\infty$ and/or $b = \infty$). A function $\varphi : (a, b) \to \mathbb{R}$ is said to be **convex** if

$$\varphi((1-\lambda)x + \lambda y) \le (1-\lambda)\varphi(x) + \lambda\varphi(y) \tag{11.40}$$

for all $x, y \in (a, b)$ and $0 \le \lambda \le 1$.

Looking at the graph of the function, the convexity condition says that, if $x < t < y$, then the point $(t, \varphi(t))$ lies below or on the line connecting the points $(x, \varphi(x))$ and $(y, \varphi(y))$ in the plane. This is equivalent to saying that

$$\frac{\varphi(t) - \varphi(s)}{t - s} \le \frac{\varphi(u) - \varphi(t)}{u - t} \tag{11.41}$$

whenever $a < s < t < u < b$. By Exercise 10.1, if $\varphi : (a, b) \to \mathbb{R}$ is convex, then it is continuous.

Theorem 11.4 (Jensen's inequality)

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) = 1$ (i.e. μ is a probability measure). Let $f \in L^1$ be a real-valued function such that $a < f(x) < b$ for all $x \in \Omega$ (here you may have $a = -\infty$ and/or $b = \infty$). If $\varphi : (a, b) \to \mathbb{R}$ is convex, then

$$\varphi\left(\int_\Omega f \, \mathrm{d}\mu\right) \leq \int_\Omega (\varphi \circ f) \, \mathrm{d}\mu. \tag{11.42}$$

Proof. Let $t = \int_\Omega f \, \mathrm{d}\mu$. Then since $a < f(x) < b$ everywhere, we have

$$\int_\Omega a \, \mathrm{d}\mu < \int_\Omega f \, \mathrm{d}\mu < \int_\Omega b \, \mathrm{d}\mu \implies a < t < b, \tag{11.43}$$

since $\mu(\Omega) = 1$. Let

$$\beta = \sup_{s \in (a,t)} \frac{\varphi(t) - \varphi(s)}{t - s}. \tag{11.44}$$

Then since $\frac{\varphi(t)-\varphi(s)}{t-s} \leq \frac{\varphi(u)-\varphi(t)}{u-t}$ for every $a < s < t < u < b$, taking supremum on the LHS, we have

$$\beta \leq \frac{\varphi(u) - \varphi(t)}{u - t}, \qquad \text{for every } u \in (t, b). \tag{11.45}$$

Now we claim that for any $a < s < b$ we have

$$\varphi(s) \geq \varphi(t) + \beta(s - t). \tag{11.46}$$

Indeed, it's true for $s = t$. If $s < t$,

$$\frac{\varphi(t) - \varphi(s)}{t - s} \leq \beta, \tag{11.47}$$

since β is the supremum. Simplifying gives

$$\varphi(t) - \varphi(s) \leq \beta\,(t - s) \implies \varphi(s) \geq \varphi(t) + \beta(s - t). \tag{11.48}$$

If $s > t$, then by (11.45),

$$\beta \leq \frac{\varphi(s) - \varphi(t)}{s - t} \implies \varphi(s) \geq \varphi(t) + \beta(s - t). \tag{11.49}$$

Hence (11.46) holds! Now, taking $s = f\,(x)$, we have

$$\varphi(f(x)) - \varphi(t) - \beta(f(x) - t) \geq 0 \tag{11.50}$$

for every $x \in \Omega$. Since φ is continuous, $\varphi \circ f$ is measurable. Integrating both sides of (11.50),

$$\int_\Omega (\varphi \circ f) \, \mathrm{d}\mu - \varphi(t) \int_\Omega 1 \, \mathrm{d}\mu - \beta\left(\int_\Omega f \, \mathrm{d}\mu - t\right) \geq 0. \tag{11.51}$$

Since $t = \int_\Omega f \, d\mu$, and $\mu(\Omega) = 1$, we have

$$\int_\Omega (\varphi \circ f) \, d\mu - \varphi\left(\int_\Omega f \, d\mu\right) \geq 0, \tag{11.52}$$

as desired. ■

Corollary 11.5 (Jensen's inequality for concave functions)

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) = 1$ (i.e. μ is a probability measure). Let $f \in L^1$ be a real-valued function such that $a < f(x) < b$ for all $x \in \Omega$ (here you may have $a = -\infty$ and/or $b = \infty$). If $\varphi : (a, b) \to \mathbb{R}$ is concave (i.e. $-\varphi$ is convex), then

$$\varphi\left(\int_\Omega f \, d\mu\right) \geq \int_\Omega (\varphi \circ f) \, d\mu. \tag{11.53}$$

Proof. By Jensen's inequality,

$$-\varphi\left(\int_\Omega f \, d\mu\right) \leq \int_\Omega (-\varphi \circ f) \, d\mu. \tag{11.54}$$

Now, negating both sides (and thus flipping the inequality) gives us the desired result! ■

If μ is a finite meausre, we can rescale it to make it a probability measure, by just considering the measure $\widetilde{\mu}(E) = \frac{\mu(E)}{\mu(\Omega)}$.

Corollary 11.6

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) < \infty$. Let $f \in L^1$ be a real-valued function such that $a < f(x) < b$ for all $x \in \Omega$. If $\varphi : (a, b) \to \mathbb{R}$ is convex, then

$$\varphi\left(\frac{1}{\mu(\Omega)} \int_\Omega f \, d\mu\right) \leq \frac{1}{\mu(\Omega)} \int_\Omega (\varphi \circ f) \, d\mu. \tag{11.55}$$

We call a function **strictly convex** if we have strict equality in (11.40) for all $x, y \in (a, b)$. Then for strictly convex functions, we have a variant of Jensen's inequality where the inequality is strict.

Corollary 11.7 (Jensen's inequality for stricly convex functions)

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) = 1$ (i.e. μ is a probability measure). Let $f \in L^1$ be a real-valued function such that $a < f(x) < b$ for all $x \in \Omega$. If $\varphi : (a, b) \to \mathbb{R}$ is ***strictly*** convex and f is not a constant a.e., then

$$\varphi\left(\int_\Omega f \, d\mu\right) < \int_\Omega (\varphi \circ f) \, d\mu. \tag{11.56}$$

This can be traced back from Jensen's inequality by just considering the equality case, and ruling it out since f is not constant.

Corollary 11.8 (Jensen's inequality for stricly concave functions)

Let (Ω, Σ, μ) be a measure space with $\mu(\Omega) = 1$ (i.e. μ is a probability measure). Let $f \in L^1$ be a real-valued function such that $a < f(x) < b$ for all $x \in \Omega$. If $\varphi : (a, b) \to \mathbb{R}$ is ***strictly*** concave and f is not a constant a.e., then

$$\varphi\left(\int_\Omega f \, \mathrm{d}\mu\right) > \int_\Omega (\varphi \circ f) \, \mathrm{d}\mu. \tag{11.57}$$

§12. 30 September 2026

Today we review inner product spaces and Hilbert spaces and prove the existence of orthogonal projections onto closed subspaces.

§12.i. Inner products and norms

Throughout, vector spaces are over $\mathbb{C}$. (Over $\mathbb{R}$ the same statements and proofs hold, with all complex conjugations removed and Re dropped.)

Definition 12.1 (Inner product). Let H be a vector space over $\mathbb{C}$. An **inner product** on H is a function $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$ such that for all $x, y, z \in H$ and $\lambda \in \mathbb{C}$:

1. (conjugate symmetry) $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
2. (linearity in the first variable) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$;
3. (positive definiteness) $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Note that $\langle x, x \rangle$ is real by (1), so (3) makes sense. By (1) and (2), the inner product is conjugate-linear in the second variable:

$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle , \qquad \langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle . \tag{12.1}$$

Thus a constant λ comes out of the first variable as λ and out of the second variable as $\overline{\lambda}$.

Example 12.1

For $f, g \in L^2(\Omega)$ (complex-valued), set

$$\langle f, g \rangle = \int_\Omega f \overline{g} \, d\mu. \tag{12.2}$$

This satisfies all the axioms. (Over $\mathbb{R}$ one simply takes $\int_\Omega fg \, d\mu$.)

Every inner product induces a norm.

$$\|x\| = \sqrt{\langle x, x \rangle}, \qquad x \in H. \tag{12.3}$$

Theorem 12.1 (Cauchy–Schwarz inequality)

Let H be a vector space with an inner product. Then for all $x, y \in H$,

$$|\langle x, y \rangle| \leq \|x\| \, \|y\| . \tag{12.4}$$

The inequality is named after Hermann Schwarz (no "t"), the German mathematician. The Frenchman Laurent Schwartz, with a "t", is the one of Schwartz space and Schwartz distributions.

Proof. Choose $\omega \in \mathbb{C}$ with $|\omega| = 1$ such that

$$\omega \langle x, y\rangle = |\langle x, y\rangle|, \tag{12.5}$$

(if $\langle x, y\rangle = 0$ any ω works). Then

$$\langle \omega x, y\rangle = \omega \langle x, y\rangle = |\langle x, y\rangle| \in \mathbb{R}, \tag{12.6}$$

and $\|\omega x\| = \|x\|$. For every real t, positivity of the inner product gives

$$0 \le \|\omega x + ty\|^2 = \langle \omega x, \omega x\rangle + t\langle \omega x, y\rangle + t\langle y, \omega x\rangle + t^2 \langle y, y\rangle \tag{12.7}$$

$$= \|x\|^2 + 2t \operatorname{Re}\langle \omega x, y\rangle + t^2 \|y\|^2 \tag{12.8}$$

$$= \|x\|^2 + 2t|\langle x, y\rangle| + t^2 \|y\|^2. \tag{12.9}$$

Here we used $\langle y, \omega x\rangle = \overline{\langle \omega x, y\rangle}$. The right-hand side is a quadratic polynomial in the real variable t which is nonnegative for all $t \in \mathbb{R}$. If $y = 0$ the claim is trivial. If $y \neq 0$, the polynomial $at^2 + bt + c$ with $a = \|y\|^2 > 0$, $b = 2|\langle x, y\rangle|$, $c = \|x\|^2$ is nonnegative everywhere, so its discriminant is nonpositive:

$$4|\langle x, y\rangle|^2 - 4\|x\|^2 \|y\|^2 \le 0. \tag{12.10}$$

Hence $|\langle x, y\rangle|^2 \le \|x\|^2 \|y\|^2$, and taking square roots gives the claim. ∎

Proposition 12.2

Let H be a vector space with an inner product. Then $\|x\| = \sqrt{\langle x, x\rangle}$ is a norm on H.

Proof. *Positive Definiteness.* $\|x\| = 0$ if and only if $\langle x, x\rangle = 0$ if and only if $x = 0$.

Homogeneity. For $\lambda \in \mathbb{C}$,

$$\|\lambda x\|^2 = \langle \lambda x, \lambda x\rangle = \lambda\overline{\lambda}\langle x, x\rangle = |\lambda|^2 \|x\|^2, \tag{12.11}$$

so $\|\lambda x\| = |\lambda| \|x\|$.

Triangle inequality. Expanding and using Cauchy–Schwarz inequality,

$$\|x + y\|^2 = \|x\|^2 + \langle x, y\rangle + \langle y, x\rangle + \|y\|^2 \tag{12.12}$$

$$= \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 \tag{12.13}$$

$$\le \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \tag{12.14}$$

$$\le \|x\|^2 + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2. \tag{12.15}$$

Taking square roots gives $\|x+y\| \le \|x\| + \|y\|$. ∎

Definition 12.2 (Hilbert space)**.** A **Hilbert space** is a vector space H over $\mathbb{C}$ with an inner product which is complete with respect to the induced norm $\|x\| = \sqrt{\langle x, x\rangle}$. In other words, it is a Banach space whose norm comes from an inner product.

Proposition 12.3 (Parallelogram law)

Let H be an inner product space. For all $x, y \in H$,

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 . \tag{12.16}$$

Proof. Expanding as above,

$$\|x+y\|^2 = \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 , \tag{12.17}$$
$$\|x-y\|^2 = \|x\|^2 - 2\operatorname{Re}\langle x, y\rangle + \|y\|^2 . \tag{12.18}$$

Adding the two identities, the $\pm 2\operatorname{Re}\langle x, y\rangle$ terms cancel. ∎

Remark 12.1. Conversely, a norm satisfying the parallelogram law comes from an inner product, given by the polarization identity

$$\langle x, y\rangle = \frac{1}{4}\sum_{k=0}^{3} i^k \left\|x + i^k y\right\|^2 . \tag{12.19}$$

§12.ii. Examples

Example 12.2

For any measure space (Ω, Σ, μ), $L^2(\Omega)$ with

$$\langle f, g\rangle = \int_\Omega f\overline{g}\,\mathrm{d}\mu \tag{12.20}$$

is a Hilbert space: completeness was shown earlier, and the norm

$$\|f\|_2 = \left(\int_\Omega |f|^2\,\mathrm{d}\mu\right)^{\frac{1}{2}} \tag{12.21}$$

is the one induced by the inner product. In particular, $\mathbb{C}^d$ is a Hilbert space.

Take $\Omega = \mathbb{N}$, with Σ the set of all subsets of $\mathbb{N}$ and μ the counting measure. Then $L^2(\mathbb{N}) = \ell^2$ is the space of sequences $a = (a_1, a_2, \dots)$ with $\sum_n |a_n|^2 < \infty$ (a function $\mathbb{N} \to \mathbb{C}$ is just a sequence), with

$$\langle a, b\rangle = \sum_{n=1}^{\infty} a_n \overline{b_n}, \qquad \|a\| = \left(\sum_{n=1}^{\infty} |a_n|^2\right)^{\frac{1}{2}} . \tag{12.22}$$

This is an infinite-dimensional analogue of $\mathbb{C}^d$. It is the prototype: every separable Hilbert space is isomorphic to ℓ^2, as will be seen in a functional analysis course.

Remark 12.2 (The unit ball is not compact)**.** In ℓ^2, let e_k be the sequence with 1 in position k and 0 elsewhere. Then $\|e_k\| = 1$ and $\|e_j - e_k\| = \sqrt{2}$ for $j \neq k$, so (e_k) has no convergent subsequence. Since ℓ^2 is a metric space, compactness is equivalent to sequential compactness, so the closed unit ball of ℓ^2 is not compact. The same holds in any infinite-dimensional Hilbert space.

§12.iii. Closed subspaces and orthogonal complements

A subspace $M \subseteq H$ is a subset which is itself a vector space (closed under sums and scalar multiplication). It is **closed** if it is closed in the norm topology, i.e. $M = \overline{M}$.

In finite dimensions every subspace is closed. In infinite dimensions this fails.

Example 12.3

Let $c_{00} \subseteq \ell^2$ be the set of sequences with only finitely many nonzero entries. It is a subspace, and it contains every e_j. Any $a \in \ell^2$ is the limit of its truncations

$$(a_1, \dots, a_N, 0, 0, \dots) \in c_{00}, \tag{12.23}$$

because

$$\sum_{n>N} |a_n|^2 \to 0. \tag{12.24}$$

So

$$\overline{c_{00}} = \ell^2 \neq c_{00}. \tag{12.25}$$

for instance the sequence $a_n = \frac{1}{n}$ is in ℓ^2, but not in c_{00}. Thus c_{00} is not closed.

Let $\ell^1 \subseteq \ell^2$ be the summable sequences. Indeed, if $\sum |a_n| < \infty$, then $|a_n| < 1$ for all large n, so $|a_n|^2 \leq |a_n|$ eventually and

$$\sum_{n=1}^{\infty} |a_n|^2 < \infty. \tag{12.26}$$

It is a subspace containing all e_j, so it is dense in ℓ^2, but it is not all of ℓ^2 since the sequence $a_n = \frac{1}{n}$ is in ℓ^2, but not in ℓ^1. Thus ℓ^1 is not closed in ℓ^2.

Definition 12.3 (Orthogonality)**.** Two vectors $x, y \in H$ are **orthogonal**, written $x \perp y$, if $\langle x, y \rangle = 0$. For a subspace $M \subseteq H$, the **orthogonal complement** is

$$M^{\perp} = \{x \in H : \ \langle x, m \rangle = 0 \text{ for all } m \in M\}. \tag{12.27}$$

Lemma 12.4

$M^\perp$ is a closed subspace of H, and $M \cap M^\perp = \{0\}$.

Proof. For fixed $m \in M$, the map $x \mapsto \langle x, m\rangle$ is linear and continuous, since by Cauchy–Schwarz inequality

$$\left|\langle x, m\rangle - \langle x', m\rangle\right| \le \left\|x - x'\right\| \|m\|. \tag{12.28}$$

Hence $\{x : \langle x, m\rangle = 0\}$ is a closed subspace (a kernel of a continuous linear map), and

$$M^\perp = \bigcap_{m \in M} \{x : \langle x, m\rangle = 0\} \tag{12.29}$$

is an intersection of closed subspaces, hence a closed subspace. (Equivalently: if $x_k \to x$ with $x_k \in M^\perp$, one can pass to the limit inside the inner product.) If $z \in M \cap M^\perp$, then $\langle z, z\rangle = 0$, so $z = 0$. ∎

§12.iv. The projection theorem

Theorem 12.5 (Orthogonal projection)

Let H be a Hilbert space, $M \subseteq H$ a closed subspace, and $x \in H$.

1. There exists a unique $Px \in M$ such that
$$\|x - Px\| = \operatorname{dist}(x, M) = \inf_{y \in M} \|x - y\|. \tag{12.30}$$
2. There is a unique decomposition $x = y + w$ with $y \in M$ and $w \in M^\perp$. Moreover $y = Px$; we write $w = Qx = x - Px$.
3. (Pythagoras) $\|x\|^2 = \|Px\|^2 + \|Qx\|^2$.
4. The maps $P : H \to H$ and $Q : H \to H$ are linear.

Geometrically, Px is the foot of perpendicular from x to M, i.e. the point of M closest to x.

Proof. 1. Let $\delta = \operatorname{dist}(x, M)$ and choose $z_n \in M$ with

$$\|x - z_n\| \le \delta + \frac{1}{n}, \text{ or equivalently, } \|x - z_n\| \to \delta. \tag{12.31}$$

(If $\delta = 0$ then $x \in \overline{M} = M$ because M is closed, and we may simply take $Px = x$; the argument below also covers this case.) We claim that (z_n) is Cauchy. Apply the Parallelogram law to $a = x - z_n$ and $b = x - z_m$. Since $a + b = 2\left(x - \frac{z_n + z_m}{2}\right)$ and $a - b = z_m - z_n$, we get

$$4\left\|x - \frac{z_n + z_m}{2}\right\|^2 + \|z_n - z_m\|^2 = 2\|x - z_n\|^2 + 2\|x - z_m\|^2. \tag{12.32}$$

Since M is a subspace, $\dfrac{z_n+z_m}{2}\in M$, hence $\left\|x-\dfrac{z_n+z_m}{2}\right\|\geq\delta$. Therefore

$$\|z_n-z_m\|^2\leq 2\|x-z_n\|^2+2\|x-z_m\|^2-4\delta^2\longrightarrow 0 \tag{12.33}$$

as $n,m\to\infty$. So (z_n) is Cauchy. As H is complete, $z_n\to z$ for some $z\in H$, and $z\in M$ because M is closed. By continuity of the norm,

$$\|x-z\|=\lim_{n\to\infty}\|x-z_n\|=\delta. \tag{12.34}$$

So we set $Px=z$. Now, about the uniqueness, suppose $z,z'\in M$ both satisfy $\|x-z\|=\|x-z'\|=\delta$. We apply the Parallelogram law with $a=x-z$, $b=x-z'$

$$\left\|z-z'\right\|^2=2\|x-z\|^2+2\left\|x-z'\right\|^2-4\left\|x-\frac{z+z'}{2}\right\|^2\leq 4\delta^2-4\delta^2=0, \tag{12.35}$$

so $z=z'$.

2. First, we show the uniqueness of this decomposition. If $x=y+w=y'+w'$ with $y,y'\in M$ and $w,w'\in M^\perp$, then using Lemma 12.4,

$$y-y'=w'-w\in M\cap M^\perp=\{0\}. \tag{12.36}$$

So $y=y'$ and $w=w'$. Now we'll show existence of this decomposition. Take $y=Px\in M$ and $w=x-Px$. We must show $w\in M^\perp$. Note $\|w\|=\delta$. Let $z\in M$ with $\|z\|=1$ and $\lambda\in\mathbb{C}$. Since $Px+\lambda z\in M$,

$$\begin{aligned}\delta^2&\leq\|x-(Px+\lambda z)\|^2=\|w-\lambda z\|^2 &(12.37)\\ &=\|w\|^2-\overline{\lambda}\langle w,z\rangle-\lambda\langle z,w\rangle+|\lambda|^2\|z\|^2 &(12.38)\\ &=\delta^2-2\operatorname{Re}(\overline{\lambda}\langle w,z\rangle)+|\lambda|^2. &(12.39)\end{aligned}$$

Hence

$$2\operatorname{Re}(\overline{\lambda}\langle w,z\rangle)\leq|\lambda|^2\qquad\text{for all }\lambda\in\mathbb{C}. \tag{12.40}$$

Choose $\lambda=t\langle w,z\rangle$ with $t>0$ real. Then

$$\overline{\lambda}\langle w,z\rangle=t|\langle w,z\rangle|^2 \tag{12.41}$$

and $|\lambda|^2=t^2|\langle w,z\rangle|^2$, so

$$2t|\langle w,z\rangle|^2\leq t^2|\langle w,z\rangle|^2\implies 2|\langle w,z\rangle|^2\leq t|\langle w,z\rangle|^2. \tag{12.42}$$

Letting $t\to 0^+$ gives $\langle w,z\rangle=0$. This holds for every unit vector $z\in M$, hence (by scaling) for every $z\in M$. So $w\in M^\perp$.

3. Since $Px \in M$ and $Qx \in M^\perp$, we have $\langle Px, Qx\rangle = 0$, so

$$\|x\|^2 = \|Px + Qx\|^2 = \|Px\|^2 + \|Qx\|^2 + 2\,\mathrm{Re}\,\langle Px, Qx\rangle = \|Px\|^2 + \|Qx\|^2\,. \tag{12.43}$$

4. Let $x, x' \in H$ and $\alpha, \beta \in \mathbb{C}$. Then

$$\alpha x + \beta x' = (\alpha Px + \beta Px') + (\alpha Qx + \beta Qx'), \tag{12.44}$$

where the first summand lies in M and the second in $M^\perp$, since both are subspaces. By uniqueness in (2), $P(\alpha x + \beta x') = \alpha Px + \beta Px'$. Hence P is linear. Finally $Q = I - P$ is linear as well.

■

Remark 12.3. Existence and uniqueness in (1) used the completeness, so the argument does not work in a general normed space. For instance, in $\mathbb{R}^2$ with the norm

$$\|(u, v)\|_\infty = \max(|u|\,, |v|), \tag{12.45}$$

let

$$M = \{(0, y) : y \in \mathbb{R}\} \tag{12.46}$$

and $x = (1, 0)$. Then

$$\|(1, 0) - (0, y)\|_\infty = \max(1, |y|) \tag{12.47}$$

which equals $\mathrm{dist}(x, M) = 1$ for every $|y| \le 1$. Hence, the nearest point is not unique.

Tomorrow we will prove the Riesz representation theorem for Hilbert spaces.

§13. 1 October 2026

A *Hilbert space* H is a vector space over $\mathbb{C}$ with an inner product $\langle \cdot, \cdot \rangle$ (linear in the first variable, conjugate-linear in the second) which is complete for the induced norm $\|x\| = \sqrt{\langle x, x \rangle}$. The basic example is $L^2(\Omega, \Sigma, \mu)$ with $\langle f, g \rangle = \int_\Omega f \overline{g}\, d\mu$.

Last time we proved the projection theorem: if $M \subset H$ is a closed subspace and $x \in H$, there is a unique $Px \in M$ realizing the distance from x to M, and $Qx = x - Px$ belongs to $M^\perp$, the set of vectors orthogonal to every element of M.

§13.i. When are L^p spaces Hilbert spaces

Remark 13.1. A natural question arises: is L^2 the only L^p which is a Hilbert space? Essentially yes, because the parallelogram law fails for $p \neq 2$ for most of the measure spaces. The concrete statement is:

> Suppose (Ω, Σ, μ) is a measure space that contains two disjoint sets A, B with $0 < \mu(A), \mu(B) < \infty$. Then L^p is **NOT** a Hilbert space for $p \neq 2$.

To prove this, we choose

$$f = \frac{1}{\mu(A)^{\frac{1}{p}}} \chi_A, \quad g = \frac{1}{\mu(B)^{\frac{1}{p}}} \chi_B. \tag{13.1}$$

Then for every p,

$$\|f\|_p = \|g\|_p = 1. \tag{13.2}$$

Since f, g have disjoint supports, $|f + g| = |f - g|$, so that for $1 \leq p < \infty$,

$$\|f + g\|_p^p = \|f - g\|_p^p = \int_\Omega \frac{1}{\mu(A)} \chi_A \, \mathrm{d}\mu + \int_\Omega \frac{1}{\mu(B)} \chi_B \, \mathrm{d}\mu = 2. \tag{13.3}$$

Hence,

$$\|f + g\|_p^2 + \|f - g\|_p^2 = 2 \cdot 2^{\frac{2}{p}} = 2^{1 + \frac{2}{p}}, \quad 2\|f\|_p^2 + 2\|g\|_p^2 = 4, \tag{13.4}$$

and these two are equal only when $p = 2$. For $p = \infty$,

$$\|f + g\|_\infty^2 + \|f - g\|_\infty^2 = 2, \quad 2\|f\|_\infty^2 + 2\|g\|_\infty^2 = 4. \tag{13.5}$$

Hence, for *well-behaved* measure spaces (i.e. measure spaces with two disjoint finite-measure sets), L^2 is the only L^p that is a Hilbert space.

Remark 13.2. Having two disjoint finite-measure sets is not an unreasonable condition to ask. In fact, the only measure spaces where this doesn't hold are:

1. Measure spaces where every nonempty set has infinite measure. But in this case, $L^p = \{0\}$. Because if $|f| > 0$ on a nonempty measurable set E, then $\int_\Omega |f|^p \, \mathrm{d}\mu = \infty$.

2. There is exactly one "component" of finite measure. Suppose there is a measurable set A with $0 < \mu(A) < \infty$. Then any set B disjoint from A will have either $\mu(B) = 0$ or $\mu(B) = \infty$. If we have $B \subseteq A$ is measurable, then we must have $\mu(B) = 0$ or $\mu(B) = \mu(A)$ (otherwise B and $A - B$ will be two disjoint finite-measure sets). In this case, we claim that

$$L^p = \{c \cdot \chi_A \ : \ c \in \mathbb{C}\}, \tag{13.6}$$

i.e. L^p is 1-dimensional. Indeed, for $f \in L^p$, first we show that $f = 0$ a.e. on A^c. Indeed, if

$$B_n = \left\{ x \in A^c \ : |f(x)| \geq \frac{1}{n} \right\}. \tag{13.7}$$

Since B_n is disjoint from A, its measure is either 0 or infinite. If it's infinite, then

$$\int_\Omega |f|^p \, \mathrm{d}\mu \geq \int_{B_n} |f|^p \ \mathrm{d}\mu \geq \frac{1}{n^p} \mu(B_n) = \infty, \tag{13.8}$$

which contradicts with $f \in L^p$. Hence, $\mu(B_n) = 0$ for all n. Therefore,

$$\mu(\{x \in A^c : |f(x)| > 0\}) = \mu\left(\bigcup_{n=1}^{\infty} B_n \right) \leq \sum_{n=1}^{\infty} \mu(B_n) = 0. \tag{13.9}$$

We assume f take real values. For the complex case, we can just work with $\operatorname{Re} f$ and $\operatorname{Im} f$ separately. For $t \in \mathbb{R}$, define

$$E_t = \{x \in A : f(x) \leq t\}. \tag{13.10}$$

Each $E_t \subseteq A$, so $\mu(E_t) = 0$ or $\mu(E_t) = \mu(A)$. Also, for $t < t'$, $E_t \subseteq E_{t'}$, so $\mu(E_t) \subseteq \mu(E_{t'})$. Now, as $t \to -\infty$, $\mu(E_t) \to 0$, and as $t \to \infty$, $\mu(E_t) \to \mu(A)$. We pick

$$c = \inf \{t \in \mathbb{R} : \mu(E_t) = \mu(A)\}. \tag{13.11}$$

The first order of business is to show that it's a finite real number. If it's not, suppose it's $-\infty$, then $\mu(E_t) = \mu(A)$ holds for every $t \in \mathbb{R}$. If $M = \int_\Omega |f|^p \, \mathrm{d}\mu$, choose

$$t = -\left(\frac{2M}{\mu(A)} \right)^{\frac{1}{p}}. \tag{13.12}$$

By hypothesis, $f(x) \leq -\left(\frac{2M}{\mu(A)}\right)^{\frac{1}{p}}$ almost everywhere on A (since $\mu(E_t) = \mu(A)$),

$$M = \int_\Omega |f|^p \, \mathrm{d}\mu \geq \int_A \left(\frac{2M}{\mu(A)} \right) \mathrm{d}\mu = 2M, \tag{13.13}$$

implies M must be 0. But this just means $f = 0$.

On the other hand, if $c = \infty$, that means there is no t for which $\mu(E_t) = \mu(A)$ happens,

i.e. $\mu(E_t) = 0$ for every t. Therefore, $\mu(A - E_t) = \mu(A)$, i.e.

$$\mu(\{x \in A : f(x) > t\}) = \mu(A), \tag{13.14}$$

for all t. We choose

$$t = \left(\frac{2M}{\mu(A)}\right)^{\frac{1}{p}}. \tag{13.15}$$

Since $f(x) > t$ almost everywhere on A (since $\mu(E_t) = 0$),

$$M = \int_\Omega |f|^p \, \mathrm{d}\mu \geq \int_A \left(\frac{2M}{\mu(A)}\right) \mathrm{d}\mu = 2M, \tag{13.16}$$

implies M must be 0. But this just means $f = 0$.

So for nonzero $f \in L^p$, c is a finite real number. We claim that $f = c \cdot \chi_A$. For $t < c$, $\mu(E_t) = 0$. Therefore,

$$\mu(\{x \in A : f(x) < c\}) = \mu\left(\bigcup_{n=1}^{\infty} \left\{x \in A : f(x) \leq c - \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{\infty} \mu\left(E_{c-\frac{1}{n}}\right) = 0. \tag{13.17}$$

On the other hand, for $t > c$, $\mu(E_t) = \mu(A)$. Hence,

$$\mu(\{x \in A : f(x) > c\}) = \mu\left(\bigcup_{n=1}^{\infty} \left\{x \in A : f(x) > c + \frac{1}{n}\right\}\right) \leq \sum_{n=1}^{\infty} \mu\left(A - E_{c+\frac{1}{n}}\right) = 0. \tag{13.18}$$

Therefore, $f = c$ almost everywhere on A, hence $f = c \cdot \chi_A$. Thus, we have proved that $\dim L^p = 1$ in this case.

Of course, for 1-dimensional normed spaces, the norm is always induced by an inner product, so this is not particularly an interesting scenario.

§13.ii. Bounded linear functionals and the dual space

Definition 13.1. Let E be a normed vector space over $\mathbb{C}$ (not necessarily complete). A **functional** on E is a function $L : E \to \mathbb{C}$.

Definition 13.2 (Linear and bounded functionals)**.** A functional $L : E \to \mathbb{C}$ is **linear** if

$$L(x + y) = L(x) + L(y) \text{ and } L(\lambda x) = \lambda L(x) \tag{13.19}$$

for all $x, y \in E$ and $\lambda \in \mathbb{C}$. It is **bounded** if there exists a constant $C \geq 0$ such that

$$|L(x)| \leq C \|x\| \qquad \text{for all } x \in E. \tag{13.20}$$

Equivalently, L maps the unit ball to a bounded set, or $\sup_{\|x\| \leq 1} |L(x)| < \infty$. In that case

we define

$$\|L\| = \sup_{\|x\|\leq 1} |L(x)| = \sup_{\|x\|=1} |L(x)|. \tag{13.21}$$

The two suprema agree by dilation: if $0 < \|x\| < 1$, then $\frac{x}{\|x\|}$ has norm 1 and

$$\left|L\left(\frac{x}{\|x\|}\right)\right| = \frac{|L(x)|}{\|x\|} \geq |L(x)|, \tag{13.22}$$

so points in the interior of the ball never beat points on the sphere. Directly from the definition,

$$|L(x)| \leq \|L\| \|x\|, \tag{13.23}$$

for all x, and $\|L\|$ is the smallest admissible constant C.

Definition 13.3 (Dual space)**.** The *dual space* E^* of E is the set of all bounded linear functionals on E.

Remark 13.3. In finite dimensions E^* is the usual dual space of linear algebra, but in infinite dimensions the space of all linear functionals is strictly larger than the space of bounded ones. The word "dual" here always refers to E^*.

E^* is a vector space: for $L_1, L_2 \in E^*$ and $\lambda \in \mathbb{C}$, set

$$(L_1 + L_2)(x) = L_1(x) + L_2(x) \tag{13.24}$$

$$(\lambda L)(x) = \lambda L(x). \tag{13.25}$$

These are again bounded linear functionals, as we'll see in the next proposition.

Proposition 13.1

$\|\cdot\|$ is a norm on E^*.

Proof. If $\|L\| = 0$, then $L = 0$ on the unit ball, hence everywhere by dilation, so L is the zero functional. For $\lambda \in \mathbb{C}$,

$$\|\lambda L\| = \sup_{\|x\|=1} |\lambda L(x)| = |\lambda| \sup_{\|x\|=1} |L(x)| = |\lambda| \|L\|. \tag{13.26}$$

Finally, for $\|x\| = 1$,

$$|(L_1 + L_2)(x)| \leq |L_1(x)| + |L_2(x)| \leq \|L_1\| + \|L_2\|, \tag{13.27}$$

and taking the supremum over x gives $\|L_1 + L_2\| \leq \|L_1\| + \|L_2\|$. ∎

Proposition 13.2

E^* is complete, even if E is not.

Proof. Let $L_1, L_2, \ldots$ be a Cauchy sequence in E^*, i.e.

$$\|L_m - L_n\| \to 0 \text{ as } m, n \to \infty. \tag{13.28}$$

Now, for every $x \in E$,

$$|L_m(x) - L_n(x)| \leq \|L_m - L_n\| \, \|x\| \to 0 \text{ as } m, n \to \infty. \tag{13.29}$$

Therefore, $L_n(x)$ is a Cauchy sequence of complex numbers. By the completeness of $\mathbb{C}$, it converges. So we can define a functional $L : E \to \mathbb{C}$

$$L(x) = \lim_{n\to\infty} L_n(x) \text{ for } x \in E. \tag{13.30}$$

L is linear by the linearity of limit and linearity of L_n's:

$$L(\alpha x + y) = \lim_{n\to\infty} L_n(\alpha x + y) = \alpha \lim_{n\to\infty} L_n(x) + \lim_{n\to\infty} L_n(y) = \alpha L(x) + L(y). \tag{13.31}$$

Now, we need to show that $L_n \to L$ in the operator norm. Given $\varepsilon > 0$ choose N large enough so that $\|L_m - L_n\| < \varepsilon$ for $m, n \geq N$.

$$|L(x) - L_n(x)| \leq |L(x) - L_m(x)| + |L_m(x) - L_n(x)| \tag{13.32}$$

Taking $m \to \infty$, we have

$$|L(x) - L_n(x)| \leq |L_m(x) - L_n(x)| \leq \varepsilon \|x\|. \tag{13.33}$$

Therefore, $\|L - L_n\| \leq \varepsilon$. Finally, $L \in E^*$, since

$$L = L_n + (L - L_n), \tag{13.34}$$

and L_n has a bounded norm, and so does $L - L_n$. ■

Proposition 13.3

A linear functional $L : E \to \mathbb{C}$ is bounded if and only if it is continuous.

Proof. Suppose L is bounded, say $|L(x)| \leq C\|x\|$. Then for all x, x_0, by linearity,

$$|L(x) - L(x_0)| = |L(x - x_0)| \leq C\|x - x_0\|, \tag{13.35}$$

so L is continuous (even Lipschitz).

Conversely, suppose L is continuous. Then it is continuous at 0, so there is $\delta > 0$ such that

$$\|y\| < \delta \text{ implies } |L(y)| < 1 \tag{13.36}$$

If $\|x\| \le 1$, then

$$\left\|\frac{\delta x}{2}\right\| < \delta \implies \left|L\left(\frac{\delta x}{2}\right)\right| < 1 \implies |L(x)| < \frac{2}{\delta}. \tag{13.37}$$

Hence L maps the unit ball into a bounded set, so L is bounded. ■

§13.iii. The Riesz representation theorem

Let H be a Hilbert space. Every vector $v \in H$ defines a functional

$$L_v(x) = \langle x, v\rangle, \qquad x \in H. \tag{13.38}$$

It is linear since the inner product is linear in the first variable. It is bounded by the Cauchy–Schwarz inequality:

$$|L_v(x)| \le \|x\|\,\|v\| \tag{13.39}$$

so $\|L_v\| \le \|v\|$. In fact $\|L_v\| = \|v\|$: if $v \neq 0$, take $x = \frac{v}{\|v\|}$, which has norm 1, and

$$L_v(x) = \frac{\langle v, v\rangle}{\|v\|} = \|v\|. \tag{13.40}$$

Thus every $v \in H$ gives an element $L_v \in H^*$ with the same norm. The following theorem says that there are no others.

Theorem 13.4 (Riesz representation theorem)

Let H be a Hilbert space and $L \in H^*$. Then there exists a unique $v \in H$ such that

$$L(x) = \langle x, v\rangle \qquad \text{for all } x \in H. \tag{13.41}$$

Moreover $\|L\| = \|v\|$.

Proof. If $L = 0$, take $v = 0$. Assume $L \neq 0$ and let

$$M = \operatorname{Ker} L = \{x \in H : \ L(x) = 0\}. \tag{13.42}$$

M is a subspace because L is linear, and it is closed because L is continuous (Proposition 13.3), since $M = L^{-1}\{0\}$ and $\{0\}$ is closed in $\mathbb{C}$. Since $L \neq 0$, there is $y \in H - M$. Let

$$z = \frac{y - Py}{\|y - Py\|}, \tag{13.43}$$

where P is the orthogonal projection onto M. Here $y - Py \neq 0$ since otherwise $y = Py \in M$. By the projection theorem (Theorem 12.5) $y - Py \in M^\perp$, so $z \in M^\perp$ and $\|z\| = 1$. Moreover

$L(z) \neq 0$, since otherwise $z \in M \cap M^{\perp} = \{0\}$, contradicting $\|z\| = 1$.

For $x \in H$, define

$$u = L(x)\, z - L(z)\, x. \tag{13.44}$$

Then

$$L(u) = L(x)\, L(z) - L(z)\, L(x) = 0, \tag{13.45}$$

so $u \in M$. Since $z \in M^{\perp}$, $\langle u, z \rangle = 0$. Expanding with linearity in the first variable and $\langle z, z \rangle = 1$,

$$0 = \langle u, z \rangle = L(x) \langle z, z \rangle - L(z) \langle x, z \rangle = L(x) - L(z) \langle x, z \rangle, \tag{13.46}$$

that is,

$$L(x) = L(z) \langle x, z \rangle = \left\langle x, \overline{L(z)}\, z \right\rangle, \tag{13.47}$$

where we used $\langle x, \lambda w \rangle = \overline{\lambda} \langle x, w \rangle$. So $v = \overline{L(z)}\, z$ works.

For the uniqueness, if $\langle x, v \rangle = \langle x, v' \rangle$ for all x, then

$$\left\langle x, v - v' \right\rangle = 0, \tag{13.48}$$

for all x. Taking $x = v - v'$ gives $\|v - v'\|^2 = 0$, so $v = v'$. ∎

Remark 13.4. The map $v \mapsto L_v$ is a bijective isometry from H onto H^*. Over $\mathbb{C}$ it is conjugate-linear, since $L_{\lambda v} = \overline{\lambda} L_v$. In this sense Hilbert spaces are self-dual. Since H^* is complete, this also shows again that H^* is a Hilbert space.

§13.iv. Signed measures

Up to now measures were nonnegative. We now allow measures with positive and negative weights. This is the setting for the Radon–Nikodym theorem, to be proved next time.

Definition 13.4 (Signed measure)**.** Let (Ω, Σ) be a measurable space. A *signed measure* on Ω is a function $\nu : \Sigma \to \mathbb{R}$ such that for every set $E \in \Sigma$ and every partition of E into countably many disjoint sets $E = \bigsqcup_{n=1}^{\infty} E_n$ with $E_n \in \Sigma$,

$$\nu(E) = \sum_{n=1}^{\infty} \nu(E_n), \qquad \text{and the series converges absolutely.} \tag{13.49}$$

(One can also consider complex-valued ν, with small modifications.)

Remark 13.5. In this definition ν takes only finite real values. Allowing $\nu(E) = \pm\infty$ would conflict with absolute convergence in (13.49), and we want to avoid expressions such as $\infty - \infty$. (In particular, Lebesgue measure on $\mathbb{R}$ is not a signed measure in this sense, since $\nu(\mathbb{R}) = \infty$.)

$\nu(\varnothing) = 0$: writing $\varnothing = \varnothing \sqcup \varnothing$ gives $\nu(\varnothing) = 2\nu(\varnothing)$.

Example 13.1

Let μ be a nonnegative measure and $f \in L^1(\mu)$ real-valued. Then

$$\nu(E) = \int_E f \, d\mu \tag{13.50}$$

is a signed measure. Indeed, for a partition $E = \bigsqcup_{n=1}^{\infty} E_n$, by Dominated Convergence Theorem, $\nu(E) = \sum_{n=1}^{\infty} \nu(E_n)$, and

$$\sum_n |\nu(E_n)| \le \sum_n \int_{E_n} |f| \, d\mu = \int_E |f| \, d\mu < \infty. \tag{13.51}$$

§13.v. The total variation measure

We want to majorize a signed measure ν by a nonnegative measure μ, i.e.

$$|\nu(E)| \le \mu(E) \tag{13.52}$$

for all E, in the *cheapest possible way*: μ should be smaller than any other nonnegative measure $\widetilde{\mu}$ with $|\nu(E)| \le \tilde{\mu}(E)$.

If $E = \bigsqcup_{n=1}^{\infty} E_n$ is a partition and μ is such a majorant, then

$$\mu(E) = \sum_{n=1}^{\infty} \mu(E_n) \ge \sum_{n=1}^{\infty} |\nu(E_n)|. \tag{13.53}$$

This suggests the following definition.

Definition 13.5 (Total variation). For a signed measure ν and $E \in \Sigma$, define $|\nu| : \Sigma \to [0, \infty]$

$$|\nu|(E) = \sup \left\{ \sum_{n=1}^{\infty} |\nu(E_n)| : \ E = \bigsqcup_{n=1}^{\infty} E_n \text{ is a countable partition of } E \right\}. \tag{13.54}$$

A priori $|\nu|$ could take the value $+\infty$. Taking the trivial partition $E = E \sqcup \varnothing \sqcup \varnothing \sqcup \cdots$ shows

$$|\nu(E)| \le |\nu|(E) \qquad \text{for all } E \in \Sigma, \tag{13.55}$$

so $|\nu|$ is a majorant. By the discussion above, any nonnegative measure majorizing ν which is additive is at least $|\nu|$ on every set. It remains to show that $|\nu|$ is itself a measure.

For $\nu(E) = \int_E f \, d\mu$, one expects $|\nu|(E) = \int_E |f| \, d\mu$.

Proposition 13.5

$|\nu|$ is a nonnegative measure on (Ω, Σ) (possibly with infinite values), called the **total variation measure** of ν.

Proof. Clearly $|\nu| \geq 0$, and $|\nu|(\varnothing) = 0$ since every partition of $\varnothing$ consists of empty sets and $\nu(\varnothing) = 0$. We show countable additivity: let $E = \bigsqcup_{n=1}^{\infty} E_n$; we prove

$$|\nu|(E) = \sum_{n=1}^{\infty} |\nu|(E_n). \tag{13.56}$$

Fix real numbers t_n with $0 \leq t_n < |\nu|(E_n)$ (and $t_n = 0$ if $|\nu|(E_n) = 0$). By the definition of the supremum, for each n there is a partition $E_n = \bigsqcup_j E_{n,j}$ with

$$\sum_j |\nu(E_{n,j})| \geq t_n \tag{13.57}$$

The family $\{E_{n,j}\}_{n,j}$ is a countable partition of E, so

$$|\nu|(E) \geq \sum_n \sum_j |\nu(E_{n,j})| \geq \sum_n t_n. \tag{13.58}$$

Taking the supremum over all such choices of t_n gives $|\nu|(E) \geq \sum_{n=1}^{\infty} |\nu|(E_n)$, which holds also if the right-hand side is $+\infty$.

Now we need to show the reverse inequality. Let $E = \bigsqcup_j A_j$ be any countable partition of E. For each j,

$$A_j = \bigsqcup_n (A_j \cap E_n) \tag{13.59}$$

and for each n,

$$E_n = \bigsqcup_j (A_j \cap E_n). \tag{13.60}$$

Using the additivity of ν and exchanging sums of nonnegative terms,

$$\sum_j |\nu(A_j)| = \sum_j \left| \sum_n \nu(A_j \cap E_n) \right| \leq \sum_j \sum_n |\nu(A_j \cap E_n)| \tag{13.61}$$

$$= \sum_n \sum_j |\nu(A_j \cap E_n)| \leq \sum_n |\nu|(E_n), \tag{13.62}$$

because $\{A_j \cap E_n\}_j$ is a partition of E_n. Taking the supremum over partitions $\{A_j\}$ gives $|\nu|(E) \leq \sum_n |\nu|(E_n)$. ∎

Lemma 13.6

Let ν be a signed measure and $E \in \Sigma$ with $|\nu|(E) = \infty$. Then $E = A \sqcup B$ with $A, B \in \Sigma$ such that

$$|\nu(A)| > 1, \qquad |\nu(B)| > 1, \qquad \text{and at least one of } |\nu|(A), |\nu|(B) \text{ is } \infty.$$

Proof. Let $t > 0$ be a number to be chosen. Since $|\nu|(E) = \infty > t$, there is a partition

$E = \bigsqcup_n E_n$ with

$$\sum_{n=1}^{\infty} |\nu(E_n)| > t \tag{13.63}$$

Let A_+ be the union of those E_n with $\nu(E_n) > 0$, and A_- the union of those with $\nu(E_n) < 0$. By absolute convergence and additivity,

$$\nu(A_+) = \sum_{\nu(E_n)>0} \nu(E_n), \qquad -\nu(A_-) = \sum_{\nu(E_n)<0} |\nu(E_n)|, \tag{13.64}$$

and these two sums add up to more than t. So one of them exceeds $\frac{t}{2}$; let A be the corresponding set, so that $|\nu(A)| > \frac{t}{2}$. Let $B = E - A$. Then

$$|\nu(B)| = |\nu(E) - \nu(A)| \geq |\nu(A)| - |\nu(E)| > \frac{t}{2} - |\nu(E)|. \tag{13.65}$$

Here $|\nu(E)|$ is a finite number, since ν is real-valued. Choose $t > 2(|\nu(E)| + 1)$; then $|\nu(A)| > \frac{t}{2} > 1$ and $|\nu(B)| > 1$. Finally, by Proposition 13.5, $|\nu|(A) + |\nu|(B) = |\nu|(E) = \infty$, so at least one of the two is infinite. ■

Proposition 13.7

If ν is a signed measure, then $|\nu|(\Omega) < \infty$.

Proof. Suppose $|\nu|(\Omega) = \infty$. By Lemma 13.6, $\Omega = A_1 \sqcup B_1$ with

$$|\nu(A_1)| > 1, \quad |\nu(B_1)| > 1, \quad |\nu|(B_1) = \infty \tag{13.66}$$

(We rename the sets so that $|\nu|(B_1) = \infty$.) Applying the lemma again to B_1, we obtain $B_1 = A_2 \sqcup B_2$ with

$$|\nu(A_2)| > 1, \quad |\nu(B_2)| > 1, \quad |\nu|(B_2) = \infty \tag{13.67}$$

Repeating, we obtain pairwise disjoint sets $A_1, A_2, A_3, \ldots$ with $|\nu(A_n)| > 1$ for all n. Let $A = \bigsqcup_n A_n$. By the definition of a signed measure, $\sum_n \nu(A_n)$ must converge absolutely, but that's not possible here as $|\nu(A_n)| > 1$. Hence $|\nu|(\Omega) < \infty$. ■

Remark 13.6. For complex-valued ν, apply the same results to the real signed measures $\operatorname{Re}\nu$ and $\operatorname{Im}\nu$, using $|\nu|(E) \leq |\operatorname{Re}\nu|(E) + |\operatorname{Im}\nu|(E)$. Alternatively, in Lemma 13.6 one chooses among four sets (positive and negative real parts, positive and negative imaginary parts). The finiteness of $|\nu|(\Omega)$ is the measure-theoretic analogue of the notion of a function of bounded variation, which will come up later.

Define

$$\nu^+ = \frac{|\nu| + \nu}{2}, \qquad \nu^- = \frac{|\nu| - \nu}{2}. \tag{13.68}$$

Both are nonnegative, because $|\nu(E)| \leq |\nu|(E)$; both are countably additive, being combinations of countably additive set functions; and both are finite by Proposition 13.7. Thus $\nu^{\pm}$ are finite

nonnegative measures and

$$\nu = \nu^+ - \nu^-, \qquad |\nu| = \nu^+ + \nu^-. \tag{13.69}$$

This is the *Jordan decomposition* of ν: every signed measure is the difference of two finite nonnegative measures.

For $\nu(E) = \int_E f \, \mathrm{d}\mu$, it corresponds to writing $f = f^+ - f^-$ with $f^\pm \geq 0$, so

$$\nu^\pm(E) = \int_E f^\pm \, \mathrm{d}\mu. \tag{13.70}$$

§13.vi. The space of signed measures

The collection of signed measures on (Ω, Σ) is a vector space: the sum of two signed measures and any scalar multiple of a signed measure (positive or negative) is again a signed measure. By Proposition 13.7, we set

$$\|\nu\| = |\nu|(\Omega) < \infty. \tag{13.71}$$

Proposition 13.8

(13.71) defines a norm on the space of signed measures.

Proof. $\|\nu\| = 0$ implies $|\nu|(\Omega) = 0$, which implies $|\nu|(E) = 0$ for every $E \in \Sigma$, in particular, $\nu(E) = 0$ for every $E \in \Sigma$.

For $\alpha \in \mathbb{R}$, and partitions $\Omega = \bigsqcup_{n=1}^\infty E_n$,

$$\|\alpha\,\nu\| = \sup_{\text{partitions}} \sum_{n=1}^\infty |\alpha\,\nu(E_n)| = |\alpha| \sup_{\text{partitions}} \sum_{n=1}^\infty |\nu(E_n)| = |\alpha|\,\|\nu\|. \tag{13.72}$$

Now, if ν_1, ν_2 are two signed measures, then

$$|\nu_1(E) + \nu_2(E)| \leq |\nu_1(E)| + |\nu_2(E)|. \tag{13.73}$$

Therefore, for a partition $\Omega = \bigsqcup_{n=1}^\infty E_n$,

$$\sum_{n=1}^\infty |\nu_1(E_n) + \nu_2(E_n)| \leq \sum_{n=1}^\infty |\nu_1(E_n)| + \sum_{n=1}^\infty |\nu_2(E_n)|. \tag{13.74}$$

Taking supremum over all partitions, we get

$$\|\nu_1 + \nu_2\| \leq \|\nu_1\| + \|\nu_2\|. \tag{13.75}$$

■

Proposition 13.9

The space of signed measures is a Banach space with respect to the norm defined in (13.71).

Proof. Let $\{\nu_n\}_{n=1}^{\infty}$ be a Cauchy sequence of signed measures, i.e. given $\varepsilon > 0$, there exists N such that for any $m, n \geq N$,

$$\|\nu_m - \nu_n\| = |\nu_m - \nu_n|(\Omega) < \varepsilon. \tag{13.76}$$

Now, for any $E \in \Sigma$,

$$|\nu_m(E) - \nu_n(E)| \leq |\nu_m - \nu_n|(E) \leq |\nu_m - \nu_n|(\Omega) < \varepsilon. \tag{13.77}$$

Therefore, $\{\nu_n(E)\}$ is a Cauchy sequence of real numbers. By the completeness of $\mathbb{R}$, we define

$$\nu(E) = \lim_{n\to\infty} \nu_n(E). \tag{13.78}$$

We have to now show that ν is a signed measure, and $\nu_n \to \nu$ in norm. Given $E \in \Sigma$ and any partition $E = \bigsqcup_{n=1}^{\infty} E_n$, for $m, n \geq N$,

$$\sum_{i=1}^{\infty} |\nu_n(E_i) - \nu_m(E_i)| \leq |\nu_m - \nu_n|(E) \leq \|\nu_n - \nu_n\| \leq \varepsilon. \tag{13.79}$$

So the bound works for any finite sum as well:

$$\sum_{i=1}^{N} |\nu_n(E_i) - \nu_m(E_i)| \leq \varepsilon. \tag{13.80}$$

Taking $m \to \infty$, we have

$$\sum_{i=1}^{N} |\nu_n(E_i) - \nu(E_i)| \leq \varepsilon. \tag{13.81}$$

Since this holds for any N, we can take $N \to \infty$ to obtain:

$$\sum_{i=1}^{\infty} |\nu_n(E_i) - \nu(E_i)| \leq \varepsilon. \tag{13.82}$$

This can also be done using Fatou's Lemma, because

$$\sum_{i=1}^{\infty} |\nu_n(E_i) - \nu(E_i)| \leq \liminf_{m\to\infty} \sum_{i=1}^{\infty} |\nu_n(E_i) - \nu_m(E_i)| \leq \varepsilon. \tag{13.83}$$

Therefore,

$$\left|\nu(E) - \sum_{i=1}^{\infty} \nu(E_i)\right| \leq |\nu(E) - \nu_n(E)| + \left|\nu_n(E) - \sum_{i=1}^{\infty} \nu_n(E_i)\right| + \sum_{i=1}^{\infty} |\nu_n(E_i) - \nu(E_i)| \tag{13.84}$$

Since $\nu_n(E) \to \nu(E)$, the first term goes to 0 as $n \to \infty$. Since ν_n is a signed measure,

$\nu_n(E) = \sum_{i=1}^{\infty} \nu_n(E_i)$, so the middle term is 0. The last term is bounded by ε as given in (13.82). Therefore,

$$\nu(E) = \sum_{i=1}^{\infty} \nu(E_i), \tag{13.85}$$

and hence ν is a signed measure on Ω.

Now let $\Omega = \bigsqcup_{n=1}^{\infty} E_n$ be a partition of Ω. Then using (13.82), we have

$$\sum_{i=1}^{\infty} |\nu_n(E_i) - \nu(E_i)| \leq \varepsilon. \tag{13.86}$$

Taking the supremum over all partitions, we have

$$|\nu_n - \nu|(\Omega) = \sup_{\text{partitions}} \sum_{i=1}^{\infty} |\nu_n(E_i) - \nu(E_i)| \leq \varepsilon. \tag{13.87}$$

Therefore, $\|\nu_n - \nu\| \to 0$. ■

§14. 12 October 2026

Let (Ω, Σ) be a measurable space. Last time we considered **signed measures** $\nu : \Sigma \to \mathbb{R}$: they take only finite real values (never $\pm\infty$) and are countably additive, in the sense that $\nu\left(\bigsqcup_n E_n\right) = \sum_n \nu(E_n)$ with the series converging absolutely. We defined the **total variation measure**

$$|\nu|(E) = \sup\left\{\sum_n |\nu(E_n)| : \ E = \bigsqcup_n E_n \text{ a countable partition}\right\}, \tag{14.1}$$

and proved that $|\nu|$ is a nonnegative measure with $|\nu|(\Omega) < \infty$, which is called the **total variation** of ν. The Jordan decomposition is $\nu = \nu^+ - \nu^-$ with

$$\nu^{\pm} = \frac{|\nu| \pm \nu}{2} \tag{14.2}$$

finite nonnegative measures.

Our guiding example: if μ is a nonnegative measure and $f \in L^1(\mu)$ is real-valued (at least locally integrable, so that the values are finite), then

$$\nu(E) = \int_E f \, \mathrm{d}\mu \tag{14.3}$$

is a signed measure, and $|\nu|(E) = \int_E |f| \, \mathrm{d}\mu$. The goal of today is to understand when a measure is of this form.

§14.i. Absolute continuity and singularity

Definition 14.1 (Absolute continuity)**.** Let ν be a signed measure and μ a nonnegative measure on (Ω, Σ). We say that ν is **absolutely continuous** with respect to μ, and write $\nu \ll \mu$, if

$$\mu(A) = 0 \implies \nu(A) = 0 \qquad \text{for all } A \in \Sigma. \tag{14.4}$$

In other words, if μ does not see a set, then ν cannot see it either.

Definition 14.2 (Support)**.** A (signed or nonnegative) measure ν is **supported** on a set $A \in \Sigma$ if

$$\nu(E) = \nu(E \cap A) \qquad \text{for all } E \in \Sigma. \tag{14.5}$$

This is equivalent to

$$\nu(B) = 0 \text{ for every } B \in \Sigma \text{ with } B \cap A = \varnothing \tag{14.6}$$

Indeed, if ν is supported on A and $B \cap A = \varnothing$, then

$$\nu(B) = \nu(B \cap A) = \nu(\varnothing) = 0. \tag{14.7}$$

Conversely,

$$\nu(E) = \nu(E \cap A) + \nu(E - A) = \nu(E \cap A) \tag{14.8}$$

if ν vanishes on subsets of $\Omega - A$. For a nonnegative measure this is also equivalent to $\nu(\Omega - A) = 0$.

Note that if ν is supported on A and $A \subseteq A'$, then ν is also supported on A', because $B \cap A' = \varnothing$ implies $B \cap A = \varnothing$, so we have

$$\nu(B) = 0. \tag{14.9}$$

So ν vanishes on subsets of $\Omega - A'$, i.e. ν is supported on A'.

Definition 14.3 (Mutual singularity)**.** Two measures ν_1, ν_2 are **mutually singular**, written $\nu_1 \perp \nu_2$, if there are disjoint sets $A_1, A_2 \in \Sigma$ with $A_1 \cap A_2 = \varnothing$ such that ν_1 is supported on A_1 and ν_2 is supported on A_2.

So ν_1 only sees what happens in A_1, ν_2 only sees what happens in A_2, and these sets do not overlap. (The notation $\perp$ is only a notation; it is not related to an inner product on the space of measures.)

Proposition 14.1 (Basic properties)

Let ν, ν_1, ν_2 be signed measures and μ a nonnegative measure.

1. If ν is supported on A, then $|\nu|$ is supported on A.
2. If $\nu_1 \perp \nu_2$, then $|\nu_1| \perp |\nu_2|$.
3. If $\nu_1 \perp \mu$ and $\nu_2 \perp \mu$, then $\nu_1 + \nu_2 \perp \mu$.
4. If $\nu_1 \ll \mu$ and $\nu_2 \ll \mu$, then $\nu_1 + \nu_2 \ll \mu$.
5. If $\nu \ll \mu$, then $|\nu| \ll \mu$.
6. If $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$, then $\nu_1 \perp \nu_2$.
7. If $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu = 0$.

The same holds with ν_1, ν_2 replaced by $-\nu_1, -\nu_2$, since $-\nu$ is supported on the same sets as ν and is absolutely continuous with respect to the same measures.

Proof. 1. Let $B \in \Sigma$ with $B \cap A = \varnothing$ and let $B = \bigsqcup_{n=1}^{\infty} B_n$ be a partition. Each B_n is disjoint from A, so $\nu(B_n) = 0$ and

$$\sum_{n=1}^{\infty} |\nu(B_n)| = 0 \tag{14.10}$$

Taking the supremum over all partitions, $|\nu|(B) = 0$.

2. If ν_i is supported on A_i with $A_1 \cap A_2 = \varnothing$, then by 1 so is $|\nu_i|$.

3. Let ν_i be supported on A_i and μ on B_i, with $A_i \cap B_i = \varnothing$, for $i = 1, 2$. Then $\nu_1 + \nu_2$ is supported on $A_1 \cup A_2$: because for a set $B \cap (A_1 \cup A_2) = \varnothing$, we have

$$\varnothing = B \cap (A_1 \cup A_2) = (B \cap A_1) \cup (B \cap A_2), \tag{14.11}$$

so $B \cap A_i = \varnothing$ for $i = 1, 2$; and hence,

$$(\nu_1 + \nu_2)(B) = \nu_1(B) + \nu_2(B) = 0 + 0 = 0. \tag{14.12}$$

Furthermore, μ is supported on $B_1 \cap B_2$, since

$$\mu(\Omega - (B_1 \cap B_2)) \leq \mu(\Omega - B_1) + \mu(\Omega - B_2) = 0. \tag{14.13}$$

Moreover,

$$(A_1 \cup A_2) \cap (B_1 \cap B_2) = \varnothing \tag{14.14}$$

because $A_1 \cap B_1 = \varnothing = A_2 \cap B_2$. Therefore, $\nu_1 + \nu_2 \perp \mu$.

4. If $\mu(A) = 0$, then

$$(\nu_1 + \nu_2)(A) = \nu_1(A) + \nu_2(A) = 0. \tag{14.15}$$

5. Let $\mu(A) = 0$. In any partition $A = \bigsqcup_{n=1}^{\infty} A_n$ we have $\mu(A_n) = 0$, hence $\nu(A_n) = 0$, so every sum

$$\sum_{n=1}^{\infty} |\nu(A_n)| = 0, \tag{14.16}$$

and taking the supremum over all partitions, $|\nu|(A) = 0$.

6. Let ν_2 be supported on A_2 and μ on B, with $A_2 \cap B = \varnothing$. Then

$$\mu(\Omega - B) = 0, \tag{14.17}$$

so for every $C \subseteq \Omega - B$,

$$\nu_1(C) = 0. \tag{14.18}$$

That is, ν_1 is supported on B. Since $A_2 \cap B = \varnothing$, $\nu_1 \perp \nu_2$.

7. By 6 with $\nu_1 = \nu_2 = \nu$, we have $\nu \perp \nu$. So ν is supported on A and on A' with $A \cap A' = \varnothing$. For any $E \in \Sigma$,

$$\nu(E) = \nu(E \cap A), \tag{14.19}$$

and $E \cap A$ is disjoint from A', so

$$\nu(E \cap A) = 0. \tag{14.20}$$

Hence $\nu = 0$.

■

Remark 14.1. In general, given a signed measure ν and a nonnegative measure μ, it may happen that $\nu \ll \mu$, that $\nu \perp \mu$, or neither. The Lebesgue decomposition theorem below says that in the "neither" case ν always splits into one absolutely continuous part and one singular part.

§14.ii. A finite equivalent of a σ-finite measure

Recall that μ is **σ-finite** if $\Omega = \bigcup_{n=1}^{\infty} K_n$ with $\mu(K_n) < \infty$ for all n.

Lemma 14.2

Let μ be a σ-finite nonnegative measure. Then there exists $w \in L^1(\mu)$ with

$$0 < w(x) < 1 \qquad \text{for every } x \in \Omega. \tag{14.21}$$

Consequently $\tilde{\mu}(E) = \int_E w \, d\mu$ is a finite measure with $\tilde{\mu}(E) = 0$ if and only if $\mu(E) = 0$.

For example, for Lebesgue measure on $\mathbb{R}$ one may take

$$w(x) = \frac{1}{2(1+x^2)} \in (0,1). \tag{14.22}$$

Proof of Lemma 14.2. Write $\Omega = \bigcup_n K_n$ with $\mu(K_n) < \infty$, and put

$$w_n = \frac{1}{2^{n+1}(1+\mu(K_n))}\chi_{K_n}, \qquad w = \sum_{n=1}^{\infty} w_n. \tag{14.23}$$

Each w_n is measurable with

$$0 \le w_n \le \frac{1}{2^{n+1}} \tag{14.24}$$

so

$$w \le \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} = \frac{1}{2} < 1. \tag{14.25}$$

If $x \in \Omega$, then $x \in K_n$ for some n, so $w(x) \ge w_n(x) > 0$. By Proposition 2.8,

$$\int_\Omega w \, d\mu = \sum_{n=1}^{\infty} \int_\Omega w_n \, d\mu \tag{14.26}$$

$$= \sum_{n=1}^{\infty} \frac{\mu(K_n)}{2^{n+1}(1+\mu(K_n))} \tag{14.27}$$

$$\le \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} = \frac{1}{2} < \infty. \tag{14.28}$$

So $w \in L^1(\mu)$ and $\tilde{\mu}(\Omega) = \int_\Omega w \, d\mu < \infty$. If $\mu(E) = 0$ then $\tilde{\mu}(E) = 0$. Conversely, if $\mu(E) > 0$, write

$$E = \bigcup_{k=1}^{\infty} \left\{ x \in E : w(x) > \frac{1}{k} \right\} =: \bigcup_{k=1}^{\infty} E_k.. \tag{14.29}$$

Some set in the union has positive μ-measure, say $\mu(E_k) > 0$. Then $w > \frac{1}{k}$ on E_k, so

$$\tilde{\mu}(E) = \int_E w \, d\mu \geq \int_{E_k} w \, d\mu \geq \frac{1}{k} \mu(E_k) > 0. \tag{14.30}$$

■

Thus $\tilde{\mu}$ and μ are mutually absolutely continuous: they see the same null sets, but $\tilde{\mu}$ is finite. This is where σ-finiteness will be used.

§14.iii. The Lebesgue–Radon–Nikodym theorem

Theorem 14.3 (Lebesgue–Radon–Nikodym theorem)

Let μ be a σ-finite nonnegative measure and ν a signed measure on (Ω, Σ).

1. (Lebesgue decomposition) There exist unique signed measures ν_a, ν_s such that
$$\nu = \nu_a + \nu_s, \qquad \nu_a \ll \mu, \qquad \nu_s \perp \mu. \tag{14.31}$$
2. (Radon–Nikodym) There exists a unique (up to μ-null sets) function $h \in L^1(\mu)$ such that
$$\nu_a(E) = \int_E h \, d\mu \qquad \text{for all } E \in \Sigma. \tag{14.32}$$

In particular, if $\nu \ll \mu$, then $\nu_s = 0$ and $\nu(E) = \int_E h \, d\mu$. This shows:

> Every absolutely continuous measure is the integral of an L^1 function against the reference measure.

The function h is called the **Radon–Nikodym derivative** of ν_a with respect to μ, written

$$h =: \frac{d\nu_a}{d\mu}, \qquad \text{equivalently} \qquad d\nu_a = h \, d\mu. \tag{14.33}$$

Remark 14.2 (Uniqueness of the decomposition)**.** If $\nu = \nu_a + \nu_s = \nu_a' + \nu_s'$ are two such decompositions, then

$$\nu_a - \nu_a' = \nu_s' - \nu_s. \tag{14.34}$$

By Proposition 14.1 parts 4 and 3, the left side is absolutely continuous with respect to μ and the right side is singular with respect to μ. By Proposition 14.1 part 7 both sides vanish, so $\nu_a = \nu_a'$ and $\nu_s = \nu_s'$. So only existence needs work.

Remark 14.3 (Uniqueness of h)**.** If $h, h' \in L^1(\mu)$ both work, then

$$\int_E h \, d\mu = \int_E h' \, d\mu, \tag{14.35}$$

for all $E \in \Sigma$. Then by Proposition 3.8, $h = h'$ μ-almost everywhere, i.e. $h = h'$ in $L^1(\mu)$.

Conversely, for every $h \in L^1(\mu)$, the set function

$$E \mapsto \int_E h \, \mathrm{d}\mu \tag{14.36}$$

is a signed measure, and it is absolutely continuous with respect to μ (because if $\mu(E) = 0$ the integral vanishes). So Theorem 14.3 says that the absolutely continuous measures are *exactly* the integrals $\int_E h \, \mathrm{d}\mu$ with $h \in L^1(\mu)$.

Proof of Lebesgue–Radon–Nikodym theorem. We shall divide the proof into several steps.

Step 0: Suppose the theorem holds when ν is a finite nonnegative measure. For a general signed measure, use the Jordan decomposition $\nu = \nu^+ - \nu^-$, where $\nu^\pm$ are finite nonnegative measures. Applying the nonnegative case to each,

$$\nu^\pm = \nu_a^\pm + \nu_s^\pm, \tag{14.37}$$

with

$$\nu_a^\pm(E) = \int_E h^\pm \, \mathrm{d}\mu \tag{14.38}$$

for $h^\pm \in L^1(\mu)$. Put

$$\nu_a = \nu_a^+ - \nu_a^-, \qquad \nu_s = \nu_s^+ - \nu_s^-, \qquad h = h^+ - h^-. \tag{14.39}$$

Then $\nu = \nu_a + \nu_s$, $\nu_a \ll \mu$ (by Proposition 14.1 part 4) and $\nu_a(E) = \int_E h \, \mathrm{d}\mu$, and $\nu_s \perp \mu$ (by Proposition 14.1 part 3). So from now on ν is a **finite nonnegative measure**.

Step 1: Let w be as in Lemma 14.2 and define

$$\eta(E) = \nu(E) + \int_E w \, \mathrm{d}\mu, \qquad \text{for } E \in \Sigma. \tag{14.40}$$

This is a finite nonnegative measure, with $\nu \le \eta$. For every measurable $f \ge 0$,

$$\int_\Omega f \, \mathrm{d}\eta = \int_\Omega f \, \mathrm{d}\nu + \int_\Omega f w \, \mathrm{d}\mu. \tag{14.41}$$

(This follows from Proposition 2.13 and Proposition 2.14.) Writing $f = f^+ - f^-$, (14.41) also holds for every $f \in L^1(\eta)$ (because everything is finite).

Step 2: All functions are real-valued, and $L^2(\eta)$ is a real Hilbert space. Since η is finite,

$$L^2(\eta) \subseteq L^1(\eta) \subseteq L^1(\nu) \tag{14.42}$$

(the first containment follows from Remark 9.9, and the second one holds because $\nu \le \eta$). Define $\Lambda : L^2(\eta) \to \mathbb{R}$ by

$$\Lambda(f) = \int_\Omega f \, \mathrm{d}\nu, \qquad f \in L^2(\eta). \tag{14.43}$$

This is well defined on classes (since an η-null set is ν-null) and linear. Λ is also bounded: by (14.41) and Hölder's inequality (or Cauchy–Schwarz inequality),

$$|\Lambda(f)| \leq \int_\Omega |f| \, d\nu \leq \int_\Omega |f| \, d\eta \tag{14.44}$$

$$\leq \left(\int_\Omega 1^2 \, d\eta \right)^{1/2} \left(\int_\Omega |f|^2 \, d\eta \right)^{1/2} \tag{14.45}$$

$$= \eta(\Omega)^{1/2} \|f\|_{L^2(\eta)}. \tag{14.46}$$

By the Riesz representation theorem, there exists $g \in L^2(\eta)$ such that $\Lambda(f) = \langle f, g \rangle$ for all $f \in L^2(\eta)$, i.e.

$$\int_\Omega f \, d\nu = \int_\Omega fg \, d\eta. \tag{14.47}$$

Step 3: We'll show that $0 \leq g \leq 1$. Taking $f = \chi_E$ in (14.47), (which lies in $L^2(\eta)$ since η is finite)

$$\nu(E) = \int_E g \, d\eta, \quad \text{for all } E \in \Sigma. \tag{14.48}$$

Furthermore, we know that

$$0 \leq \nu(E) \leq \eta(E), \quad \text{for all } E \in \Sigma. \tag{14.49}$$

Let's first take $E = \{x \in \Omega \ : \ g(x) < 0\}$ If $\eta(E) > 0$, then

$$\nu(E) = \int_E g \, d\eta < 0 \tag{14.50}$$

contradicting $\nu(E) \geq 0$. So

$$\eta(\{x \in \Omega \ : \ g(x) < 0\}) = 0. \tag{14.51}$$

Now, let $E = \{x \in \Omega \ : \ g(x) > 1\}$. If $\eta(E) > 0$, then

$$\nu(E) - \eta(E) = \int_E (g - 1) \, d\eta > 0. \tag{14.52}$$

contradicting $\nu(E) \leq \eta(E)$. So

$$\eta(\{x \in \Omega \ : \ g(x) > 1\}) = 0. \tag{14.53}$$

So $0 \leq g \leq 1$ η-almost everywhere. Changing g on an η-null set does not affect (14.47), so we may assume

$$0 \leq g(x) \leq 1 \qquad \text{for every } x \in \Omega. \tag{14.54}$$

Step 4: Combine (14.47) and (14.41) (applied to $fg \in L^1(\eta)$, since g is bounded):

$$\int_\Omega f \, d\nu = \int_\Omega fg \, d\nu + \int_\Omega fgw \, d\mu. \tag{14.55}$$

Moving the first term on the right to the left (all terms are finite for $f \in L^2(\eta)$),

$$\int_\Omega f(1-g)\,\mathrm{d}\nu = \int_\Omega fgw\,\mathrm{d}\mu \qquad \text{for all } f \in L^2(\eta). \tag{14.56}$$

Define the two disjoint sets

$$A = \{x \in \Omega :\ 0 \le g(x) < 1\}, \qquad B = \{x \in \Omega :\ g(x) = 1\}, \qquad A \sqcup B = \Omega. \tag{14.57}$$

Set

$$\nu_a(E) = \nu(E \cap A), \qquad \nu_s(E) = \nu(E \cap B). \tag{14.58}$$

These are finite nonnegative measures and $\nu = \nu_a + \nu_s$, since

$$\nu(E) = \nu(E \cap (A \sqcup B)) = \nu(E \cap A) + \nu(E \cap B) = \nu_a(E) + \nu_s(E). \tag{14.59}$$

Step 5: Now, we'll show that $\nu_s \perp \mu$. Put $f = \chi_B$ in (14.56). The LHS is

$$\int_\Omega \chi_B(1-g)\,\mathrm{d}\nu = \int_B (1-g)\,\mathrm{d}\nu = 0, \tag{14.60}$$

since $g = 1$ on B. The RHS is

$$\int_\Omega \chi_B gw = \int_B w\,\mathrm{d}\mu. \tag{14.61}$$

Hence,

$$\int_B w\,\mathrm{d}\mu = 0, \tag{14.62}$$

and since $w > 0$ everywhere, $\mu(B) = 0$. Thus μ is supported on A, as

$$\mu(\Omega - A) = \mu(B) = 0, \tag{14.63}$$

while ν_s is supported on B, and $A \cap B = \varnothing$. So $\nu_s \perp \mu$.

Step 6: At this stage we'll show $\nu_a \ll \mu$. Fix $E \in \Sigma$ and $n \in \mathbb{Z}_{>0}$ and put

$$f = \left(1 + g + g^2 + \cdots + g^n\right)\chi_E, \tag{14.64}$$

which is bounded (by $n+1$) and hence lies in $L^2(\eta)$. Now, (14.56) gives

$$\int_\Omega \left(1 + g + g^2 + \cdots + g^n\right)\chi_E(1-g)\,\mathrm{d}\nu = \int_\Omega \left(1 + g + g^2 + \cdots + g^n\right)\chi_E gw\,\mathrm{d}\mu. \tag{14.65}$$

Using the fact that $(1-g)(1 + g + \cdots + g^n) = 1 - g^{n+1}$, we get

$$\int_E \left(1 - g^{n+1}\right)\mathrm{d}\nu = \int_E \left(g + g^2 + \cdots + g^{n+1}\right) w\,\mathrm{d}\mu. \tag{14.66}$$

Let's take $n \to \infty$. Since $0 \le g \le 1$, both integrands are nonnegative and nondecreasing in n, so the Monotone Convergence Theorem applies on both sides. On the left, $1 - g^{n+1} \to \chi_A$

pointwise (it tends to 1 where $g < 1$, and equals 0 where $g = 1$), so

$$\lim_{n\to\infty} \int_E \left(1 - g^{n+1}\right) \mathrm{d}\nu = \int_E \lim_{n\to\infty} \left(1 - g^{n+1}\right) \mathrm{d}\nu = \int_E \chi_A \,\mathrm{d}\nu = \nu\left(E \cap A\right) = \nu_a\left(E\right) \tag{14.67}$$

On the right, set

$$h = w \sum_{k=1}^{\infty} g^k \in [0, \infty], \tag{14.68}$$

so that

$$\lim_{n\to\infty} \int_E \left(g + g^2 + \cdots + g^{n+1}\right) w \,\mathrm{d}\mu = \int_E h \,\mathrm{d}\mu. \tag{14.69}$$

Therefore, with (14.67) and (14.69), we have

$$\nu_a\left(E\right) = \int_E h \,\mathrm{d}\mu \qquad \text{for all } E \in \Sigma. \tag{14.70}$$

Taking $E = \Omega$ gives

$$\int_\Omega h \,\mathrm{d}\mu = \nu_a\left(\Omega\right) \leq \nu\left(\Omega\right) < \infty, \tag{14.71}$$

so $h \in L^1\left(\mu\right)$; in particular $h < \infty$, μ-almost everywhere (indeed $h = \infty$ only possibly on B since $g = 1$ on B, and $\mu\left(B\right) = 0$), and redefining $h = 0$ on B gives a finite-valued representative. Finally $\nu_a \ll \mu$ because

$$\int_E h \,\mathrm{d}\mu = 0, \qquad \text{when } \mu\left(E\right) = 0. \tag{14.72}$$

This completes the proof. ■

Where the hypotheses were used: σ-finiteness of μ gave the weight w and the finite measure η; finiteness of ν made Λ bounded; the Riesz representation theorem produced g, and the Monotone Convergence Theorem turned the geometric series into the density h. The result extends to nonnegative σ-finite measures, where the density need not be integrable.

Corollary 14.4

Let ν and μ be σ-finite nonnegative measures on (Ω, Σ) (both may take the value $+\infty$). Then there exist unique nonnegative σ-finite measures ν_a, ν_s with

$$\nu = \nu_a + \nu_s, \qquad \nu_a \ll \mu, \qquad \nu_s \perp \mu, \tag{14.73}$$

and a measurable function $h : \Omega \to [0, \infty)$, unique μ-almost everywhere, such that

$$\nu_a\left(E\right) = \int_E h \,\mathrm{d}\mu \qquad \text{for all } E \in \Sigma. \tag{14.74}$$

Both sides may be $+\infty$, and h is not necessarily in $L^1\left(\mu\right)$.

Proof. Since ν and μ are σ-finite, we can write $\Omega = \bigsqcup_{n=1}^{\infty} K_n$ with $K_n \in \Sigma$ pairwise disjoint and $\nu\left(K_n\right) < \infty$. Let $\nu_n\left(E\right) = \nu\left(E \cap K_n\right)$, a finite measure. By Theorem 14.3 (the nonnegative

case, which gives $h_n \geq 0$),

$$\nu_n = \nu_{n,a} + \nu_{n,s}, \qquad \nu_{n,a}(E) = \int_E h_n \, \mathrm{d}\mu, \quad h_n \in L^1(\mu), \; h_n \geq 0. \tag{14.75}$$

Since $\nu_n(\Omega - K_n) = 0$, also $\nu_{n,a}(\Omega - K_n) = 0$, so $h_n = 0$ μ-a.e. outside K_n; we may take $h_n = 0$ outside K_n. Define

$$\nu_a = \sum_{n=1}^{\infty} \nu_{n,a}, \qquad \nu_s = \sum_{n=1}^{\infty} \nu_{n,s}, \qquad h = \sum_{n=1}^{\infty} h_n. \tag{14.76}$$

At each point only one h_n is nonzero, so h is finite everywhere, and by Monotone Convergence Theorem (actually Proposition 2.8),

$$\nu_a(E) = \int_E h \, \mathrm{d}\mu. \tag{14.77}$$

If $\mu(E) = 0$, every $\nu_{n,a}(E) = 0$, so $\nu_a \ll \mu$. If $\nu_{n,s}$ is supported on A_n and μ on B_n with $A_n \cap B_n = \varnothing$, then ν_s is supported on $A = \bigcup_n A_n$, μ is supported on $B = \bigcap_n B_n$, and $A \cap B = \varnothing$; so $\nu_s \perp \mu$. Both parts are σ-finite since $\nu_a(K_n), \nu_s(K_n) \leq \nu(K_n) < \infty$. Uniqueness follows from the finite case applied to the restrictions to each K_n. ∎

Remark 14.4. Only the absolutely continuous part ν_a is described by μ. The singular part ν_s lives on a set where μ vanishes, and in general nothing more can be said about it. For example, let μ be Lebesgue measure on $\mathbb{R}$ and $\nu = \delta_0$ the Dirac mass at 0. Then ν is supported on $\{0\}$, and μ is supported on $\mathbb{R} - \{0\}$, so $\nu \perp \mu$ and $\nu_a = 0$.

In the language of the Cantor function (to be seen later), the measure of a monotone function splits into an absolutely continuous part, a jump part (Dirac masses) and a Cantor part, but in a general measure space there is no such finer description.

Proposition 14.5

Let ν be a signed measure and μ a non-negative measure on Σ. The following are equivalent:

(i) $\nu \ll \mu$.

(ii) For every $\varepsilon > 0$ there exists $\delta > 0$ such that, for any set E, if $\mu(E) < \delta$ then $|\nu(E)| < \varepsilon$.

Proof. It is clear that (ii) implies (i). Let us prove the converse. Assume that (i) holds but (ii) does not. Then there exists an $\varepsilon > 0$ for which no $\delta > 0$ works. Then we can find sets E_n with $\mu(E_n) < 2^{-n}$ but $|\nu(E_n)| \geq \varepsilon$. Hence $|\nu|(E_n) \geq \varepsilon$ as well. Let us write

$$A_n = \bigcup_{i=n}^{\infty} E_i \quad \text{and} \quad A = \bigcap_{n=1}^{\infty} A_n.$$

Note that $A_1 \supseteq A_2 \supseteq \ldots$ and that $\mu(A_n) < 2^{-n+1}$. In this setup we have already established that $\mu(A) = 0$ (Proposition 1.4) and that

$$|\nu|(A) = \lim_{n\to\infty} |\nu|(A_n) \geq \liminf_{n\to\infty} |\nu|(E_n) \geq \varepsilon. \tag{14.78}$$

It follows that we do not have $|\nu| \ll \mu$, a contradiction to Proposition 14.1 part 5. ■

Remark 14.5. Note that in (14.78) we used the fact that $|\nu|(A_1) < \infty$, which comes from the fact that ν is a signed measure and, according to our definition, this implies that $|\nu|(\Omega) < \infty$. If ν were a non-negative measure that is not finite, the result of Proposition 14.5 would not hold (the part (i) implies (ii)). To see this, take μ being the Lebesgue measure in the interval $(0,1)$, and ν given by

$$\nu(E) = \int_E t^{-1}\,\mathrm{d}t$$

for $E \subset (0,1)$ Lebesgue measurable. Then $\mu(E) = 0$ implies that $\nu(E) = 0$. But Proposition 14.5(ii) doesn't hold, as

$$\nu((0,\varepsilon)) = \int_0^\varepsilon \frac{1}{t}\,\mathrm{d}t = \infty. \tag{14.79}$$

§14.iv. Bounded functionals in L^p: Riesz representation revisited

Given a measure space (Ω, Σ, μ) our goal here is to study the dual space $(L^p)^*$. In the case of L^2, that is a Hilbert space, we have already seen the duality appearing in Theorem 3.1. We are now interested in the general case $1 \leq p < \infty$.

Theorem 14.6 (Riesz representation)

Assume that $1 \leq p < \infty$ and that μ is a σ-finite measure. Let $L \in (L^p)^*$. Then there is a unique $g \in L^{p'}$ (recall that $\frac{1}{p} + \frac{1}{p'} = 1$) such that

$$L(f) = \int_\Omega fg\,\mathrm{d}\mu \tag{14.80}$$

for all $f \in L^p$. In this case:

$$\|L\|_{(L^p)^*} = \|g\|_{p'}. \tag{14.81}$$

Remark 14.6. In other words, the space $(L^p)^*$ is isometrically isomorphic to $L^{p'}$. Once we establish (14.80), we have seen that (14.81) follows from Proposition 11.1. We will review part of this argument below.

Proof. The uniqueness is clear. Let us prove the existence.

We first consider the case in which $\mu(\Omega) < \infty$. For any measurable set E, define

$$\nu(E) := L(\chi_E).$$

From the linearity of L, we see that

$$\nu\left(\bigcup_{n=1}^{N} E_n\right) = \sum_{n=1}^{N} \nu\left(E_n\right)$$

if the E_n's are disjoint. To prove the countable additivity let $E = \bigcup_{n=1}^{\infty} E_n$ with all E_n's disjoint. Let $A_k = \bigcup_{n=1}^{k} E_n$ and note that

$$\left\|\chi_E - \chi_{A_k}\right\|_p = \mu\left(E - A_k\right)^{\frac{1}{p}} \to 0$$

as $k \to \infty$ (note the assumption $p < \infty$ being used here). The continuity of L now shows that $\nu\left(A_k\right) \to \nu\left(E\right)$ as we wanted. We have then established that ν is a signed measure. From the definition, if $\mu\left(E\right) = 0$ we have $\nu\left(E\right) = 0$, and hence ν is absolutely continuous with respect to μ. The Lebesgue–Radon–Nikodym theorem guarantees the existence of a function $g \in L^1\left(\mu\right)$ such that, for every measurable set E we have

$$\nu\left(E\right) = L\left(\chi_E\right) = \int_E g \, \mathrm{d}\mu = \int_\Omega \chi_E g \, \mathrm{d}\mu. \tag{14.82}$$

By linearity we have that

$$L\left(\varphi\right) = \int_\Omega \varphi g \, \mathrm{d}\mu \tag{14.83}$$

for any simple function φ. If $f \in L^\infty$, we have seen that we can construct a sequence of simple functions φ_k such that $0 \leq \left|\varphi_1\right| \leq \left|\varphi_2\right| \leq \ldots \leq \left|f\right|$ and such that $\left\|\varphi_k - f\right\|_\infty \to 0$. Since $\mu\left(\Omega\right) < \infty$ we have by Hölder's inequality that $\left\|\varphi_k - f\right\|_p \to 0$ as $k \to \infty$. Applying the continuity of L on the left-hand side of (14.83) and the Dominated Convergence Theorem on the right-hand side of (14.83) (since $\varphi_n g$ is dominated by $\left|fg\right| \in L^1$) with this sequence of φ_k we find that

$$L\left(f\right) = \int_\Omega fg \, \mathrm{d}\mu. \tag{14.84}$$

We would like to conclude that $g \in L^{p'}$ and that (14.84) holds in fact for all $f \in L^p$. At this moment we split our argument into two cases:

Case 1. $p = 1$. In this case, from (14.82) we immediately get

$$\left|\int_E g \, \mathrm{d}\mu\right| \leq \|L\| \left\|\chi_E\right\|_1 = \|L\| \, \mu\left(E\right)$$

for every measurable set E, and this plainly leads us to $\left\|g\right\|_\infty \leq \|L\|$.

Case 2. $1 < p < \infty$. Let $E_n = \{x \in \Omega : \left|g\left(x\right)\right| \leq n\}$ and set

$$f_n = \chi_{E_n} \operatorname{sgn}\left(g\left(x\right)\right) \left|g\right|^{p'-1}.$$

Then $f_n \in L^\infty$ and we plug it in (14.84) to get

$$\int_{E_n} |g|^{p'} \, \mathrm{d}\mu = \int_\Omega f_n g \, \mathrm{d}\mu = L(f_n) \leq \|L\| \, \|f_n\|_p = \|L\| \left(\int_{E_n} |g|^{p'} \, \mathrm{d}\mu \right)^{\frac{1}{p}}$$

and therefore

$$\left(\int_\Omega \chi_{E_n} |g|^{p'} \, \mathrm{d}\mu \right)^{\frac{1}{p}'} \leq \|L\| .$$

Sending $n \to \infty$ and applying the Monotone Convergence Theorem we get that $\|g\|_{p'} \leq \|L\|$.

We have then successfully proved that $g \in L^{p'}$. It follows by Hölder's inequality that both sides of (14.84) are continuous functions on L^p. Since these agree on the dense subspace $L^\infty \subset L^p$, they must coincide for all $f \in L^p$. Of course, by Hölder's inequality we have $\|L\| \leq \|g\|_{p'}$ and hence

$$\|L\| = \|g\|_{p'} .$$

This concludes the proof in the case $\mu(\Omega) < \infty$.

If $\mu(\Omega) = \infty$ we choose a function w given by Lemma 14.2. Then $\tilde{\mu}(E) = \int_E w \, \mathrm{d}\mu$ is a finite measure on Ω and the map

$$F \mapsto w^{\frac{1}{p}} F$$

is a linear isometry from $L^p(\tilde{\mu})$ onto $L^p(\mu)$ (recall that $w(x) > 0$ for all $x \in \Omega$). Hence

$$\tilde{L}(F) := L\left(w^{\frac{1}{p}} F \right)$$

defines a bounded linear functional on $L^p(\tilde{\mu})$, with $\left\| \tilde{L} \right\|_{(L^p(\tilde{\mu}))^*} = \|L\|_{(L^p(\mu))^*}$. We have already proved the existence of a function $G \in L^{p'}(\tilde{\mu})$ such that

$$\tilde{L}(F) = \int_\Omega FG \, d\tilde{\mu}$$

for all $F \in L^p(\tilde{\mu})$. Put $g = w^{\frac{1}{p'}} G$ (if $p = 1$ then put $g = G$). Then, if $p > 1$ we have

$$\int_\Omega |g|^{p'} \, \mathrm{d}\mu = \int_\Omega |G|^{p'} \, d\tilde{\mu} = \left\| \tilde{L} \right\|^{p'} = \|L\|^{p'} .$$

If $p = 1$ we directly have

$$\|g\|_\infty = \|G\|_\infty = \left\| \tilde{L} \right\| = \|L\| .$$

In particular $g \in L^{p'}$. Since $G \, \mathrm{d}\tilde{\mu} = w^{\frac{1}{p}} g \, \mathrm{d}\mu$, we finally get

$$L(f) = \tilde{L}\left(w^{-\frac{1}{p}} f \right) = \int_\Omega \left(w^{-\frac{1}{p}} f \right) G \, d\tilde{\mu} = \int_\Omega fg \, \mathrm{d}\mu$$

for every $f \in L^p(\mu)$. The proof is now complete. ∎

§14.v. Exercises

Exercise 14.1. Let ν be a signed measure on Σ. Show that there exists a real-valued measurable function h with $|h(x)| = 1$ for all $x \in \Omega$ and such that

$$\mathrm{d}\nu = h \,\mathrm{d}\,|\nu|\,. \tag{14.85}$$

Solution. Since $|\nu(E)| \leq |\nu|(E)$ for every $E \in \Sigma$, if $|\nu|(E) = 0$, then we have $\nu(E) = 0$. Therefore, $\nu \ll |\nu|$. As a result, by Lebesgue–Radon–Nikodym theorem, there exists a unique function $h \in L^1(|\nu|)$ such that

$$\nu(E) = \int_E h \,\mathrm{d}\,|\nu|\,. \tag{14.86}$$

Now we just need to prove that $|h| = 1$ everywhere in Ω. Since (14.86) doesn't change if we alter h on a $|\nu|$-null set, it suffices to show that $|h| = 1$ $|\nu|$-a.e.

For $E \in \Sigma$, let $E = \bigsqcup_{i=1}^{\infty} E_i$ be a partition into disjoint measurable sets. Then

$$|\nu|(E) = \sup_{\text{partitions}} \sum_{i=1}^{\infty} |\nu(E_i)| = \sup_{\text{partitions}} \sum_{i=1}^{\infty} \left| \int_{E_i} h \,\mathrm{d}\,|\nu| \right| \tag{14.87}$$

$$\leq \sup_{\text{partitions}} \sum_{i=1}^{\infty} \int_{E_i} |h| \,\mathrm{d}\,|\nu| = \sup_{\text{partitions}} \int_E |h| \,\mathrm{d}\,|\nu| \tag{14.88}$$

$$= \int_E |h| \,\mathrm{d}\,|\nu| \tag{14.89}$$

Therefore,

$$\int_E (|h| - 1) \,\mathrm{d}\,|\nu| = \int_E |h| \,\mathrm{d}\,|\nu| - |\nu|(E) \geq 0, \tag{14.90}$$

so $|h| - 1 \geq 0$, i.e. $|h| \geq 1$ $|\nu|$-a.e. Let

$$E^+ = \{x \in E : h(x) \geq 0\} \text{ and } E^- = \{x \in E : h(x) < 0\}\,. \tag{14.91}$$

Then $E = E^+ \sqcup E^-$, so by the definition of total variation measure,

$$|\nu|(E) \geq \left|\nu\left(E^+\right)\right| + \left|\nu\left(E^-\right)\right|. \tag{14.92}$$

Now,

$$\left|\nu\left(E^+\right)\right| = \left| \int_{E^+} h \,\mathrm{d}\,|\nu| \right| = \int_{E^+} h \,\mathrm{d}\,|\nu| = \int_{E^+} |h| \,\mathrm{d}\,|\nu|\,, \tag{14.93}$$

$$\left|\nu\left(E^-\right)\right| = \left| \int_{E^-} h \,\mathrm{d}\,|\nu| \right| = \int_{E^-} -h \,\mathrm{d}\,|\nu| = \int_{E^-} |h| \,\mathrm{d}\,|\nu|\,. \tag{14.94}$$

After adding,

$$\left|\nu\left(E^+\right)\right| + \left|\nu\left(E^-\right)\right| = \int_{E^+} |h| \,\mathrm{d}\,|\nu| + \int_{E^-} |h| \,\mathrm{d}\,|\nu| = \int_E |h| \,\mathrm{d}\,|\nu|\,. \tag{14.95}$$

Therefore, if we combine (14.89), (14.92), (14.95), we have

$$\int_E |h| \ \mathrm{d}\,|\nu| \leq |\nu|(E) = \int_E 1 \ \mathrm{d}\,|\nu| \leq \int_E |h| \ \mathrm{d}\,|\nu|. \tag{14.96}$$

Hence, $h = 1 \ \ |\nu|$-a.e. ■

Exercise 14.2. Let μ be a non-negative measure on Σ, $g \in L^1(\mu)$ and

$$\nu(E) = \int_E g \ \mathrm{d}\mu. \tag{14.97}$$

Show that

$$|\nu|(E) = \int_E |g| \ \mathrm{d}\mu. \tag{14.98}$$

Solution. For $E \in \Sigma$, let $E = \bigsqcup_{i=1}^{\infty} E_i$ be a partition into disjoint measurable sets. Then

$$|\nu|(E) = \sup_{\text{partitions}} \sum_{i=1}^{\infty} |\nu(E_i)| = \sup_{\text{partitions}} \sum_{i=1}^{\infty} \left| \int_{E_i} g \ \mathrm{d}\mu \right| \tag{14.99}$$

$$\leq \sup_{\text{partitions}} \sum_{i=1}^{\infty} \int_{E_i} |g| \ \mathrm{d}\mu = \sup_{\text{partitions}} \int_E |g| \ \mathrm{d}\mu \tag{14.100}$$

$$= \int_E |g| \ \mathrm{d}\mu \tag{14.101}$$

Now, let

$$E^+ = \{x \in E : g(x) \geq 0\} \text{ and } E^- = \{x \in E : g(x) < 0\}. \tag{14.102}$$

Then $E = E^+ \sqcup E^-$, so by the definition of total variation measure,

$$|\nu|(E) \geq \left|\nu\left(E^+\right)\right| + \left|\nu\left(E^-\right)\right|. \tag{14.103}$$

Now,

$$\left|\nu\left(E^+\right)\right| = \left| \int_{E^+} g \ \mathrm{d}\mu \right| = \int_{E^+} g \ \mathrm{d}\mu = \int_{E^+} |g| \ \mathrm{d}\mu, \tag{14.104}$$

$$\left|\nu\left(E^-\right)\right| = \left| \int_{E^-} g \ \mathrm{d}\mu \right| = \int_{E^-} -g \ \mathrm{d}\mu = \int_{E^-} |g| \ \mathrm{d}\mu. \tag{14.105}$$

After adding,

$$\left|\nu\left(E^+\right)\right| + \left|\nu\left(E^-\right)\right| = \int_{E^+} |g| \ \mathrm{d}\mu + \int_{E^-} |g| \ \mathrm{d}\mu = \int_E |g| \ \mathrm{d}\mu. \tag{14.106}$$

Combining (14.101), (14.103), (14.106), we are done! ■

Alternate Solution. We can also solve this using Exercise 14.1. There exists a measurable function h such that

$$\nu(E) = \int_E h \ \mathrm{d}\,|\nu|. \tag{14.107}$$

Therefore,

$$|\nu|(E) = \int_E 1 \,\mathrm{d}\,|\nu| = \int_E h^2 \,\mathrm{d}\,|\nu| = \int_E h \,\mathrm{d}\nu. \tag{14.108}$$

But $\nu(E) = \int_E g \,\mathrm{d}\mu$, so

$$|\nu|(E) = \int_E h \,\mathrm{d}\nu. = \int_E gh \,\mathrm{d}\mu. \tag{14.109}$$

Now $|\nu|$ is a non-negative measure, so we must have that $gh \geq 0$ μ-almost everywhere. Furthermore, $|gh| = |g|$, so $gh \geq 0$ means $gh = |g|$ μ-almost everywhere. Therefore,

$$|\nu|(E) = \int_E gh \,\mathrm{d}\mu = \int_E |g| \,\mathrm{d}\mu. \tag{14.110}$$

■

Exercise 14.3. Let ν be a signed measure on Σ.

(i) [Hahn decomposition] There exist disjoint sets A and B with $A \cup B = \Omega$ such that

$$\nu^+(E) = \nu(E \cap A) \quad \text{and} \quad \nu^-(E) = -\nu(E \cap B) \tag{14.111}$$

for all $E \in \Sigma$.

(ii) If $\nu = \nu_1 - \nu_2$ is a decomposition of ν as a difference of two non-negative measures, then $\nu_1 \geq \nu^+$ and $\nu_2 \geq \nu^-$.

Solution. (i) By Exercise 14.1, there exists a real-valued measurable function h with $|h| = 1$ everywhere, such that

$$\nu(E) = \int_E h \,\mathrm{d}\,|\nu|. \tag{14.112}$$

Let $A = h^{-1}(\{1\})$ and $B = h^{-1}(\{-1\})$. Then $A \cup B = \Omega$ and $A \cap B = \varnothing$. Now, since $\frac{1+h}{2} = \chi_A$,

$$\begin{aligned}
\nu^+(E) &= \frac{1}{2}\left[|\nu|(E) + \nu(E)\right] && (14.113)\\
&= \frac{1}{2}\left[\int_E 1 \,\mathrm{d}\,|\nu| + \int_E h \,\mathrm{d}\,|\nu|\right] && (14.114)\\
&= \int_E \frac{1+h}{2} \,\mathrm{d}\,|\nu| = |\nu|(E \cap A) && (14.115)
\end{aligned}$$

Since $h = 1$ on $E \cap A$,

$$|\nu|(E \cap A) = \int_{E\cap A} 1 \;\,\mathrm{d}\,|\nu| = \int_{E\cap A} h \;\,\mathrm{d}\,|\nu| = \nu(E \cap A). \tag{14.116}$$

Therefore, $\nu^+(E) = \nu(E \cap A)$. Now, since $E = (E \cap A) \cup (E \cap B)$,

$$\nu(E) = \nu(E \cap A) + \nu(E \cap B). \tag{14.117}$$

So we have

$$\nu(E \cap B) = \nu(E) - \nu(E \cap A) = \nu(E) - \nu^+(E) = -\nu^-(E). \tag{14.118}$$

(ii) Since $\nu \leq \nu_1$,

$$\nu^+(E) = \nu(E \cap A) \leq \nu_1(E \cap A) \leq \nu_1(E). \tag{14.119}$$

Similarly, $-\nu \leq \nu_2$, so

$$\nu^-(E) = -\nu(E \cap B) \leq \nu_2(E \cap A) \leq \nu_2(E). \tag{14.120}$$

■

Exercise 14.4. In this problem we denote a point $x \in \mathbb{R}^2$ by $x = (x_1, x_2)$.

(i) Let $A \subset \mathbb{R}^2$ be a Borel set. Prove or disprove: the set

$$A_0 = \{x_1 \in \mathbb{R} : (x_1, 0) \in A\} \subset \mathbb{R} \tag{14.121}$$

is a Borel set of $\mathbb{R}$.

(ii) Define two measures μ_1 and μ_2 on the Borel subsets of $\mathbb{R}^2$ as follows. For a Borel set $A \subset \mathbb{R}^2$

$$\mu_1(A) = \left|A \cap \left\{x \in \mathbb{R}^2 : x_2 = 0\right\}\right|_{\mathcal{L}^1} + \int_A e^{-|x|^2}\,\mathrm{d}x, \tag{14.122}$$

and

$$\mu_2(A) = \left|A \cap \left\{x \in \mathbb{R}^2 : |x|^2 \leq 1\right\}\right|_{\mathcal{L}^2}. \tag{14.123}$$

Here $|\cdot|_{\mathcal{L}^1}$ is the one-dimensional Lebesgue measure and $|\cdot|_{\mathcal{L}^2}$ is the two-dimensional Lebesgue measure. Determine the Lebesgue–Radon–Nikodym decomposition of μ_1 with respect to μ_2.

Solution. (i) Yes, A_0 is a Borel set. Consider the map

$$\begin{aligned} \iota : \mathbb{R} &\hookrightarrow \mathbb{R}^2 \\ t &\mapsto (t, 0). \end{aligned} \tag{14.124}$$

Then ι is continuous, in particular, it is Borel measurable, i.e. preimages of Borel measurable sets are Borel measurable. Therefore, $A_0 = \iota^{-1}(A)$ is a Borel set.

(ii) Let $L = \left\{x \in \mathbb{R}^2 : x_2 = 0\right\}$ be the x-axis, and $B = \left\{x \in \mathbb{R}^2 : |x|^2 \leq 1\right\}$ be the closed unit ball. Then

$$\mu_1(A) = |A \cap L|_{\mathcal{L}^1} + \int_A e^{-|x|^2}\,\mathrm{d}x = |A \cap L|_{\mathcal{L}^1} + \int_{A \cap B^c} e^{-|x|^2}\,\mathrm{d}x + \int_{A \cap B} e^{-|x|^2}\,\mathrm{d}x. \tag{14.125}$$

We claim that

$$\mu_{1,a}(A) = \int_{A\cap B} e^{-|x|^2}\,\mathrm{d}x \text{ and } \mu_{1,s}(A) = |A\cap L|_{\mathcal{L}^1} + \int_{A\cap B^c} e^{-|x|^2}\,\mathrm{d}x. \tag{14.126}$$

First, if $\mu_2(A) = |A\cap B|_{\mathcal{L}^2} = 0$, then any Lebesgue integral over $A\cap B$ is 0, and hence,

$$\mu_{1,a} = \int_{A\cap B} e^{-|x|^2}\,\mathrm{d}x = 0. \tag{14.127}$$

Now, let $E = L\cup B^c$. We claim that $\mu_{1,s}$ is suppprted on E and μ_2 is supported on E^c. Note that they are non-negative measures.

$$\mu_2(E) \le \mu_2(L) + \mu_2(B^c) = 0 + 0 = 0, \tag{14.128}$$

since $|L|_{\mathcal{L}^2} = 0$. Also, $E^c = L^c\cap B$.

$$\mu_{1,s}(L^c\cap B) = |L^c\cap B\cap L|_{\mathcal{L}^1} + \int_{L^c\cap B\cap B^c} e^{-|x|^2}\,\mathrm{d}x = 0. \tag{14.129}$$

Hence, $\mu_{1,s}\perp\mu_2$, and (14.126) is the LRN decomposition of μ_1 with respect to μ_2.

■

Exercise 14.5. Define three measures on the Borel subsets $E\subseteq[-2,2]$ as follows:

(i) $\mu_1(E)$ is the Lebesgue measure.

(ii) $\mu_2(E)$ is the Lebesgue measure of $E\cap(0,1)$.

(iii) $\mu_3(E) = \int_E x^2\,\mathrm{d}x$.

For each pair $i\neq j$ determine the Lebesgue–Radon–Nikodym decomposition of μ_i with respect to μ_j. That is, determine a function $f_{i,j}$ defined on $[-2,2]$, and a measure $\nu_{i,j}$ defined on the Borel subsets of $[-2,2]$ such that

$$\mu_i(E) = \int_E f_{i,j}(x)\ \mathrm{d}\mu_j(x) + \nu_{i,j}(E) \quad \text{for all Borel subsets } E\subset[-2,2],$$

and such that the measures μ_j and $\nu_{i,j}$ are mutually singular.

Solution. With respect to μ_1:

- If $\mu_1(E) = 0$, then $\mu_2(E) = 0$, so $\mu_2\ll\mu_1$. So its singular component is 0, and its absolutely continuous component is itself. So

$$\mu_2(E) = \mu_1(E\cap(0,1)) = \underbrace{\int_E \chi_{(0,1)}\,\mathrm{d}\mu_1}_{\int_E f_{2,1}\,\mathrm{d}\mu_1} + \underbrace{0}_{\nu_{2,1}(E)}. \tag{14.130}$$

- If $\mu_1(E) = 0$, then any Lebesgue integral over E is also 0, so $\mu(3) = 0$. So $\mu_2 \ll \mu_1$, and thus its singular component is 0, and its absolutely continuous component is itself.

$$\mu_3(E) = \int_E x^2\,dx = \underbrace{\int_E x^2\,\mathrm{d}\mu_1}_{\int_E f_{3,1}\,\mathrm{d}\mu_1} + \underbrace{0}_{\nu_{3,1}(E)}. \tag{14.131}$$

With respect to μ_2:

- The decomposition of μ_1 is:

$$\begin{aligned}\mu_1(E) &= \mu_1(E \cap (0,1)) + \mu_1(E - (0,1)) && (14.132)\\ &= \mu_2(E) + \mu_1(E - (0,1)) && (14.133)\\ &= \underbrace{\int_E 1\,\mathrm{d}\mu_2}_{\int_E f_{1,2}\,\mathrm{d}\mu_2} + \underbrace{\mu_1(E \cap (0,1)^c)}_{\nu_{1,2}(E)}. && (14.134)\end{aligned}$$

  $\nu_{1,2}(E) = \mu_1(E \cap (0,1)^c)$ is the singular part of μ_1 with respect to μ_2, because its support is $[-2,2] - (0,1)$ and the support of μ_2 is $(0,1)$, and they are disjoint.

- The decomposition of μ_3 is:

$$\begin{aligned}\mu_3(E) &= \int_E x^2\,\mathrm{d}x && (14.135)\\ &= \int_{E\cap(0,1)} x^2\,\mathrm{d}x + \int_{E\cap(0,1)} x^2\,\mathrm{d}x && (14.136)\\ &= \underbrace{\int_E x^2\,\mathrm{d}\mu_2}_{\int_E f_{3,2}\,\mathrm{d}\mu_2} + \underbrace{\int_{E\cap(0,1)} x^2\,\mathrm{d}x}_{\nu_{3,2}(E)}. && (14.137)\end{aligned}$$

  $\nu_{3,2}(E) = \int_{E\cap(0,1)} x^2\,\mathrm{d}x$ is the singular part of μ_1 with respect to μ_2, because its support is $[-2,2] - (0,1)$ and the support of μ_2 is $(0,1)$, and they are disjoint.

With respect to μ_3:

- If $\mu_3(E) = 0$, then $\int_E x^2\,\mathrm{d}x = 0$. But $x^2 > 0$ a.e, so $\int_E x^2\,\mathrm{d}x = 0$ would imply $\mu_1(E) = 0$. Hence, $\mu_1 \ll \mu_3$, and as a result, its singular part is 0 and its absolutely continuous part is itself.

$$\mu_1(E) = \int_E 1\,\mathrm{d}x = \int_E \frac{1}{x^2}\,x^2\,\mathrm{d}x = \underbrace{\int_E \frac{1}{x^2}\,\mathrm{d}\mu_3}_{\int_E f_{1,3}\mathrm{d}\mu_3} + \underbrace{0}_{\nu_{1,3}(E)}. \tag{14.138}$$

- If $\mu_3(E) = 0$, then $\int_E x^2\,\mathrm{d}x = 0$. But $x^2 > 0$ a.e, so $\int_E x^2\,\mathrm{d}x = 0$ would imply $\mu_2(E) = 0$. Hence, $\mu_2 \ll \mu_3$, and as a result, its singular part is 0 and its absolutely continuous part is

itself.

$$\mu_2(E) = \int_{E\cap(0,1)} 1\,\mathrm{d}x = \int_E \chi_{(0,1)}\frac{1}{x^2}\; x^2\,\mathrm{d}x = \underbrace{\int_E \chi_{(0,1)}\frac{1}{x^2}\;\mathrm{d}\mu_3}_{\int_E f_{2,3}\mathrm{d}\mu_3} + \underbrace{0}_{\nu_{2,3}(E)}. \tag{14.139}$$

■

§15. 14 October 2026

In this lecture, we'll discuss Fubini's theorem. The setup is as follows: Decompose $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $d = d_1 + d_2$. We write a point of $\mathbb{R}^d$ as (x, y), where $x \in \mathbb{R}^{d_1}$ is the first block of coordinates and $y \in \mathbb{R}^{d_2}$ is the second block.

Definition 15.1 (Slices)**.** Let f be a function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ (real valued; for complex valued functions treat the real and imaginary parts separately). For fixed $y \in \mathbb{R}^{d_2}$ and fixed $x \in \mathbb{R}^{d_1}$ define the **slices** $f^y : \mathbb{R}^{d_1} \to \mathbb{R}$ and $f_x : \mathbb{R}^{d_2} \to \mathbb{R}$:

$$f^y(x) = f(x, y), \qquad f_x(y) = f(x, y). \tag{15.1}$$

Similarly, for a set $E \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ define

$$E^y = \left\{ x \in \mathbb{R}^{d_1} : (x, y) \in E \right\}, \qquad E_x = \left\{ y \in \mathbb{R}^{d_2} : (x, y) \in E \right\}. \tag{15.2}$$

Note that E^y is *not* the projection of E: for fixed y it is the preimage of E under the inclusion $x \mapsto (x, y)$. In terms of characteristic functions, $(\chi_E)^y = \chi_{E^y}$.

Remark 15.1. If $E \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is Lebesgue measurable, this does *not* imply that every slice E^y is measurable. For example, let $W \subseteq \mathbb{R}^{d_1}$ be a non-measurable set and put

$$E = W \times \{0\} \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}. \tag{15.3}$$

Since E is contained in the hyperplane $\mathbb{R}^{d_1} \times \{0\}$, it has measure zero in $\mathbb{R}^d$, hence it is measurable. All slices E^y with $y \neq 0$ are empty, but $E^0 = W$ is not measurable. So the slices are measurable only for *almost every* y, which is exactly what the theorem below asserts.

Remark 15.2. If E is Borel measurable on $\mathbb{R}^d$, then all slices E^y are, in fact, Borel measurable. Because, if we take the inclusion map

$$\begin{aligned} \iota_y : \mathbb{R}^{d_1} &\hookrightarrow \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \\ x &\mapsto (x, y), \end{aligned} \tag{15.4}$$

then ι_y is continuous, and $E^y = \iota_y^{-1}(E)$. If E is a Borel set, since ι_y is continuous, preimages of Borel sets are Borel, so $E^y = \iota_y^{-1}(E)$ is Borel measurable.

Similarly, if $f : \mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ is Borel measurable, then one can show that the slice $f^y : \mathbb{R}^{d_1} \to \mathbb{R}$ is also Borel measurable, since

$$f^y(x) = f(x, y) = f(\iota_y(x)) = (f \circ \iota_y)(x), \tag{15.5}$$

so we have $f^y = f \circ \iota_y$ is the composition of two Borel measurable functions, and hence f^y is Borel measurable.

§15.i. Fubini's theorem

Theorem 15.1 (Fubini's theorem)

Let f be an integrable function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then:

(i) for almost every $y \in \mathbb{R}^{d_2}$, the slice f^y is measurable and integrable on $\mathbb{R}^{d_1}$;

(ii) the function $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\,\mathrm{d}x$, which is defined for almost every y (and set equal to 0 on the exceptional null set), is integrable on $\mathbb{R}^{d_2}$;

(iii) we have

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f^y(x)\,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^d} f \,\mathrm{d}m. \tag{15.6}$$

Remark 15.3. The statement is symmetric in x and y; the same holds with the roles of x and y exchanged, so in particular

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)\,\mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^d} f \,\mathrm{d}m = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x. \tag{15.7}$$

The moral of the theorem is that if f is integrable, you may integrate first in $\mathbb{R}^{d_1}$ and then in $\mathbb{R}^{d_2}$, or the other way round, or simply integrate over $\mathbb{R}^d$; all give the same answer. In other words, we can switch the order of integration, as long as the function is integrable, i.e. $\int |f| < \infty$.

Proof of Fubini's theorem. The proof strategy is as follows: let $\mathcal{F}$ be the class of integrable functions f on $\mathbb{R}^d$ satisfying (i), (ii) and (iii) of Fubini's theorem. We want to show that $\mathcal{F}$ is the class of *all* integrable functions. A priori $\mathcal{F}$ could be small, but it is not empty: it contains the zero function. We show in several steps that $\mathcal{F}$ is closed under certain operations and contains certain functions, until we reach all integrable functions.

Lemma 15.2

Finite linear combinations of functions in $\mathcal{F}$ are in $\mathcal{F}$.

Proof. Let $f, g \in \mathcal{F}$. For f there is a null set $E_1 \subseteq \mathbb{R}^{d_2}$ such that f^y is measurable and integrable for $y \notin E_1$; for g there is a null set E_2 with the same property. Then $E_1 \cup E_2$ is a null set, and for $y \notin E_1 \cup E_2$ the slice

$$(f+g)^y = f^y + g^y \tag{15.8}$$

is measurable and integrable. The same argument applies to scalar multiples since

$$(cf)^y = c\, f^y. \tag{15.9}$$

Moreover, (ii) and (iii) hold for $f+g$ and cf by linearity of the integral, because they hold for f and for g. Hence $f+g, cf \in \mathcal{F}$, and by induction every finite linear combination of elements of $\mathcal{F}$ lies in $\mathcal{F}$. ∎

Lemma 15.3

Let (f_k) be a sequence in $\mathcal{F}$ converging pointwise and monotonically (increasing or decreasing) to a function f, and assume f is integrable. Then $f \in \mathcal{F}$.

Proof. Replacing f_k by $-f_k$ we may assume that f_k increases to f. Replacing f_k by $f_k - f_1$,

$$f_k - f_1 \to f - f_1. \tag{15.10}$$

So we may assume that $f_k \geq 0$ for all k. Indeed, if we prove $f - f_1 \in \mathcal{F}$, then since $f_1 \in \mathcal{F}$, Lemma 15.2 gives $f \in \mathcal{F}$. So assume $0 \leq f_1 \leq f_2 \leq \cdots \to f$.

The two ideas to use are the Monotone Convergence Theorem and the fact that a countable union of null sets is null.

For each k let $E_k \subseteq \mathbb{R}^{d_2}$ be the null set outside of which f_k^y is measurable and integrable, and let $E = \bigcup_{k=1}^{\infty} E_k$. Then E is a null set, and for $y \notin E$ all the slices f_k^y are integrable (in particular measurable).

For $y \notin E$ and every x,

$$\lim_{k\to\infty} f_k^y(x) = \lim_{k\to\infty} f_k(x,y) = f(x,y) = f^y(x). \tag{15.11}$$

Thus f^y is a pointwise limit of measurable functions, hence measurable. Now we'll show that f^y is integrable for $y \notin E$. For each k, by property (iii) for f_k,

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f_k^y(x)\, \mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^d} f_k \, \mathrm{d}m. \tag{15.12}$$

Put $g_k(y) = \int_{\mathbb{R}^{d_1}} f_k^y(x)\, \mathrm{d}x$, which is an integrable function of y by (ii). For $y \notin E$, the functions f_k^y are nonnegative and increase to f^y, so by the Monotone Convergence Theorem

$$\lim_{k\to\infty} g_k(y) = g(y) := \int_{\mathbb{R}^{d_1}} f^y(x)\, \mathrm{d}x \in [0,\infty]. \tag{15.13}$$

At this point $g(y)$ might be $+\infty$. Now apply the Monotone Convergence Theorem once more, in $\mathbb{R}^{d_2}$, to the increasing nonnegative functions g_k, and then in $\mathbb{R}^d$ to $f_k \to f$. Letting $k \to \infty$ in (15.12) we get

$$\int_{\mathbb{R}^{d_2}} g(y)\, \mathrm{d}y = \lim_{k\to\infty} \int_{\mathbb{R}^{d_2}} g_k(y)\, \mathrm{d}y = \lim_{k\to\infty} \int_{\mathbb{R}^d} f_k \, \mathrm{d}m = \int_{\mathbb{R}^d} f \, \mathrm{d}m. \tag{15.14}$$

Since f is integrable by assumption, the RHS is finite. Therefore,

$$\int_{\mathbb{R}^{d_2}} g(y)\, \mathrm{d}y < \infty, \tag{15.15}$$

so $g(y) < \infty$ for almost every y, i.e.

$$\int_{\mathbb{R}^{d_1}} f^y(x)\,\mathrm{d}x < \infty \tag{15.16}$$

for almost every y. So the slice f^y is integrable for a.e. y. This proves (i). The function g is integrable on $\mathbb{R}^{d_2}$, which is (ii). Finally, (15.14) gives

$$\int_{\mathbb{R}^d} f\,\mathrm{d}m = \int_{\mathbb{R}^{d_2}} g(y)\,\mathrm{d}y = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f^y(x)\,\mathrm{d}x \right) \mathrm{d}y, \tag{15.17}$$

which is exactly (iii) for f. Hence $f \in \mathcal{F}$. ■

At this stage we do not yet know which functions lie in $\mathcal{F}$, but it is closed under finite linear combinations and monotone limits. Now we show that it contains many functions, starting with characteristic functions.

Lemma 15.4

If E is a countable intersection of open sets (a G_δ set) of finite measure, then $\chi_E \in \mathcal{F}$.

Proof. We do it in several steps.

Part 1: Let $E = Q = Q_1 \times Q_2$ be an open cube, where $Q_1 \subseteq \mathbb{R}^{d_1}$ and $Q_2 \subseteq \mathbb{R}^{d_2}$ are open cubes. Then

$$\int_{\mathbb{R}^d} \chi_Q\,\mathrm{d}m = m(Q) = |Q| = |Q_1|\,|Q_2|, \tag{15.18}$$

the product of the sides. For every y, the slice Q^y is

$$Q^y = \begin{cases} \varnothing & \text{if } y \notin Q_2, \\ Q_1 & \text{if } y \in Q_2. \end{cases} \tag{15.19}$$

Hence $\chi_Q^y = \chi_{Q^y}$ is integrable for every y, the function $y \mapsto \int \chi_Q^y\,\mathrm{d}x = |Q_1|\,\chi_{Q_2}(y)$ is integrable on $\mathbb{R}^{d_2}$, and

$$\int_{\mathbb{R}^{d_2}} |Q_1|\,\chi_{Q_2}(y)\,\mathrm{d}y = |Q_1|\,|Q_2|. \tag{15.20}$$

Hence $\chi_Q \in \mathcal{F}$. The same computation works for closed cubes.

Part 2: Let E be a subset of the boundary of a cube $Q \subseteq \mathbb{R}^d$. Then E is a null set, and for almost every $y \in \mathbb{R}^{d_2}$ the slice E^y has measure zero in $\mathbb{R}^{d_1}$. Indeed,

$$\partial(Q_1 \times Q_2) = \left(\partial Q_1 \times \overline{Q_2}\right) \cup \left(\overline{Q_1} \times \partial Q_2\right), \tag{15.21}$$

So we have

$$[\partial(Q_1 \times Q_2)]^y = \begin{cases} \partial Q_1 & \text{if } y \in Q_2^\circ, \\ \overline{Q_1} & \text{if } y \in \partial Q_2, \\ \varnothing & \text{if } y \notin Q_2. \end{cases} \tag{15.22}$$

∂Q_2 is a null-set in $\mathbb{R}^{d_2}$, so for almost every $y \in \mathbb{R}^{d_2}$, $[\partial (Q_1 \times Q_2)]^y$ is measurable with measure 0 in $\mathbb{R}^{d_1}$. By completeness of Lebesgue measure, the slices E^y are measurable with measure zero for a.e. y, and

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} \chi_E^y(x)\, \mathrm{d}x \right) \mathrm{d}y = \int_{\mathbb{R}^{d_2}} m\left(E^y\right)\, \mathrm{d}y = 0 = m(E) = \int_{\mathbb{R}^d} \chi_E. \tag{15.23}$$

Thus $\chi_E \in \mathcal{F}$.

Part 3: If $E = \bigcup_{j=1}^N Q_j$ is a finite union of almost disjoint closed cubes, then $\chi_E \in \mathcal{F}$. Indeed, χ_E equals a finite linear combination of the χ_{Q_j} up to terms supported on the overlaps, which are contained in boundaries of cubes. By Lemma 15.2, and Part 1 and Part 2, we have that $\chi_E \in \mathcal{F}$.

Part 4: Let $O \subseteq \mathbb{R}^d$ be open with $m(O) < \infty$. Write $O = \bigcup_{k=1}^\infty Q_k$ as a countable union of almost disjoint closed cubes, and let

$$O_n = \bigcup_{k=1}^n Q_k. \tag{15.24}$$

Then $\chi_{O_n} \in \mathcal{F}$ by Part 3, and $\chi_{O_n} \to \chi_O$ pointwise, monotonically increasingly. Since χ_O is integrable (finite measure), Lemma 15.3 gives $\chi_O \in \mathcal{F}$.

Part 5: Let E be a G_δ set of finite measure. Then

$$E = \bigcap_{k=1}^\infty O_k \tag{15.25}$$

with each O_k open. Since E has finite measure, we may arrange that every O_k has finite measure: choose an open set $O \supseteq E$ with $m(O - E) < 1$ (so $m(O) < \infty$) and replace O_k by $O_k \cap O$. Next make the sequence decreasing by setting

$$\widetilde{O}_n = \bigcap_{k=1}^n O_k. \tag{15.26}$$

Each $\widetilde{O}_n$ is open (a finite intersection of open sets) of finite measure, and

$$\widetilde{O}_1 \supseteq \widetilde{O}_2 \supseteq \cdots \tag{15.27}$$

with $\bigcap_n \widetilde{O}_n = E$. By Part 4, $\chi_{\widetilde{O}_n} \in \mathcal{F}$, and $\chi_{\widetilde{O}_n} \to \chi_E$ monotonically decreasingly. Since χ_E is integrable, Lemma 15.3 gives $\chi_E \in \mathcal{F}$. ■

Lemma 15.5

If E has measure zero, then $\chi_E \in \mathcal{F}$.

Proof. Let E be a null set. Then by Proposition 6.7, there exists a a G_δ set G containing E,

and $m(G - E) = 0$ and hence

$$m(G) = m(E) + m(G - E) = 0. \tag{15.28}$$

By Lemma 15.4, $\chi_G \in \mathcal{F}$. Since $\int_{\mathbb{R}^d} \chi_G = m(G) = 0$, property (iii) says

$$\int_{\mathbb{R}^{d_2}} m(G^y)\, \mathrm{d}y = 0, \tag{15.29}$$

so $m(G^y) = 0$ for almost every y. Now $E^y \subseteq G^y$, and Lebesgue measure is complete, so for almost every y the slice E^y is measurable with measure zero. Therefore χ_E^y is integrable for a.e. y, $y \mapsto m(E^y)$ vanishes a.e. (hence is integrable), and both sides of (15.6) are equal to 0. So $\chi_E \in \mathcal{F}$. ∎

(Here completeness of Lebesgue measure is used. For general measures one needs to be careful if the measure is not complete.)

Lemma 15.6

If E is measurable with $m(E) < \infty$, then $\chi_E \in \mathcal{F}$.

Proof. By Proposition 6.7, there exists a a G_δ set G containing E, and $m(G - E) = 0$.

$$\chi_E = \chi_G - \chi_{G-E}. \tag{15.30}$$

By Lemma 15.4, $\chi_G \in \mathcal{F}$; by Lemma 15.5, $\chi_{G-E} \in \mathcal{F}$; and by Lemma 15.2, $\chi_E \in \mathcal{F}$. ∎

Now, we'll complete the proof by establishing that every integrable function belongs to $\mathcal{F}$.

Let f be integrable and write

$$f = f^+ - f^-, \tag{15.31}$$

where $f^\pm \geq 0$ are integrable. By Lemma 15.2 it suffices to treat $f \geq 0$.

A nonnegative integrable function is the pointwise increasing limit of simple functions φ_k, and each φ_k is a finite linear combination of characteristic functions of measurable sets of finite measure. By Lemma 15.6 and Lemma 15.2, $\varphi_k \in \mathcal{F}$. Since $\varphi_k \to f$ monotonically increasingly, and f is integrable, Lemma 15.3 gives $f \in \mathcal{F}$.

Hence $\mathcal{F}$ is the class of all integrable functions, which proves Fubini's theorem. ∎

Remark 15.4. This is the usual "class of functions" strategy: prove the property for characteristic functions (here: cubes, then G_δ sets, open sets, measurable sets), then for simple functions, then pass to limits. The hardest part is the case of characteristic functions of measurable sets of finite measure, which is why it is divided into several steps.

§15.ii. Tonelli's theorem

If the function is *nonnegative*, the integrability hypothesis can be dropped.

Theorem 15.7 (Tonelli's theorem)

Let f be a nonnegative measurable function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then:

(i) for almost every $y \in \mathbb{R}^{d_2}$, the slice f^y is measurable on $\mathbb{R}^{d_1}$;

(ii) the function $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\,dx$, is measurable on $\mathbb{R}^{d_2}$;

(iii) we have

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f^y(x)\,dx \right) dy = \int_{\mathbb{R}^d} f\,dm, \tag{15.32}$$

where both sides may equal $+\infty$.

In other words: for nonnegative functions you may interchange the order of integration freely (both sides are simultaneously finite or infinite). For functions with sign, you must first check that the absolute value has finite integral, i.e. that $f \in L^1$; then Fubini's theorem lets you switch the order.

Proof. The basic idea here is to truncate. For each $k \in \mathbb{Z}_{>0}$ let

$$f_k(x,y) = \begin{cases} f(x,y) & \text{if } |(x,y)| \le k \text{ and } f(x,y) \le k, \\ 0 & \text{otherwise.} \end{cases} \tag{15.33}$$

Then each f_k is integrable. By Fubini's theorem, for each k, there exists a set $E_k \subseteq \mathbb{R}^{d_2}$ of measure zero such that f_k^y is integrable for $y \notin E_k$. Let $E = \bigcup_k E_k$. Then $E \subseteq \mathbb{R}^{d_2}$ still has measure zero and f_k^y is integrable (and hence measurable) for each k and $y \notin E$. For $y \notin E$, since $f_k^y \to f^y$ monotonically increasing we find that f^y is measurable and, by the Monotone Convergence Theorem,

$$\lim_{k\to\infty} \int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx = \int_{\mathbb{R}^{d_1}} f(x,y)\,dx. \tag{15.34}$$

Still by part (iii) of Fubini's theorem each map $y \mapsto \int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx$ is integrable (and hence measurable) for $y \notin E$, and hence $y \mapsto \int_{\mathbb{R}^{d_1}} f(x,y)\,dx$ is also measurable. Another application of the Monotone Convergence Theorem yields

$$\lim_{k\to\infty} \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx \right) dy = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)\,dx \right) dy. \tag{15.35}$$

On the other hand, by part (iii) of Fubini's theorem we know that

$$\int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx \right) dy = \int_{\mathbb{R}^d} f_k(x,y), \tag{15.36}$$

and by Monotone Convergence Theorem,

$$\lim_{k\to\infty}\int_{\mathbb{R}^d} f_k(x,y) = \int_{\mathbb{R}^d} f(x,y)\,. \tag{15.37}$$

By (15.35), (15.36) and (15.37), we find that

$$\int_{\mathbb{R}^{d_2}}\left(\int_{\mathbb{R}^{d_1}} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y = \int_{\mathbb{R}^d} f(x,y)\,, \tag{15.38}$$

as desired. ■

Remark 15.5 (General measures)**.** The result extends to general σ-finite measure spaces, but then one has to construct the product σ-algebra and the product measure.An important special case is a counting measure, where Fubini–Tonelli says that one may interchange sums and integrals (and double sums of nonnegative terms). One may also transfer results from $\mathbb{R}$ to sums by considering functions that are constant on intervals $[n, n+1)$.

Corollary 15.8

If E is a measurable set in $\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}$, then for almost all $y\in\mathbb{R}^{d_2}$, the slice E^y is measurable in $\mathbb{R}^{d_1}$. Moreover the function $y\mapsto m(E^y)$ is a measurable function and

$$m(E) = \int_{\mathbb{R}^{d_2}} m(E^y)\,\mathrm{d}y. \tag{15.39}$$

Proof. This is just applying Tonelli's theorem on $f=\chi_E$. ■

Proposition 15.9

If $E_1\subseteq\mathbb{R}^{d_1}$ and $E_2\subseteq\mathbb{R}^{d_2}$ then,

$$m^*(E_1\times E_2)\le m^*(E_1)\,m^*(E_2)\,, \tag{15.40}$$

with the understanding that if one of the sets E_j's has exterior measure zero then $m^*(E_1\times E_2)=0$.

Proof. Let $\varepsilon>0$. By definition, we can find cubes $\{Q_k\}_{k=1}^\infty$ in $\mathbb{R}^{d_1}$ and $\{Q'_l\}_{l=1}^\infty$ in $\mathbb{R}^{d_2}$ such that

$$E_1\subseteq\bigcup_{k=1}^\infty Q_k,\quad\text{and}\quad E_2\subseteq\bigcup_{l=1}^\infty Q'_l \tag{15.41}$$

and

$$\sum_{k=1}^\infty |Q_k|\le m^*(E_1)+\varepsilon\quad\text{and}\quad\sum_{l=1}^\infty |Q'_l|\le m^*(E_2)+\varepsilon. \tag{15.42}$$

Since $E_1 \times E_2 \subseteq \bigcup_{k,l=1}^{\infty} (Q_k \times Q'_l)$, the sub-additivity of the exterior measure yields

$$m^* (E_1 \times E_2) \leq \sum_{k,l=1}^{\infty} |Q_k \times Q'_l| = \left(\sum_{k=1}^{\infty} |Q_k| \right) \left(\sum_{l=1}^{\infty} |Q'_l| \right) \tag{15.43}$$

$$\leq (m^* (E_1) + \varepsilon) (m^* (E_2) + \varepsilon) . \tag{15.44}$$

If neither E_1 nor E_2 has exterior measure 0, then from (15.44), taking $\varepsilon \to 0^+$,

$$m^* (E_1 \times E_2) \leq m^* (E_1) m^* (E_2) . \tag{15.45}$$

If for instance $m^* (E_1) = 0$, consider for each positive integer j the set

$$E_2^j = E_2 \cap \left\{ y \in \mathbb{R}^{d_2} : |y| \leq j \right\} \tag{15.46}$$

Then, by the above argument, we find that $m^* \left(E_1 \times E_2^j \right) = 0$. Since $\left(E_1 \times E_2^j \right) \to (E_1 \times E_2)$ as $j \to \infty$, we conclude that $m^* (E_1 \times E_2) = 0$. ■

Proposition 15.10

If $E_1 \subseteq \mathbb{R}^{d_1}$ and $E_2 \subseteq \mathbb{R}^{d_2}$ are measurable sets, then $E_1 \times E_2 \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is a measurable set and

$$m (E_1 \times E_2) = m (E_1) m (E_2) , \tag{15.47}$$

with the understanding that if one of the sets E_j's has measure zero then $m (E_1 \times E_2) = 0$.

Proof. It suffices to prove that $E = E_1 \times E_2$ is measurable, because then (15.47) follows from Corollary 15.8. Since each set E_j is measurable, there exist sets $G_j \subseteq \mathbb{R}^{d_j}$ of type G_δ, with $G_j \supseteq E_j$ and

$$m^* (G_j - E_j) = 0 \tag{15.48}$$

for each $j = 1, 2$ (by Proposition 6.7). Clearly, $G = G_1 \times G_2$ is measurable in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ and

$$(G_1 \times G_2) - (E_1 \times E_2) \subseteq ((G_1 - E_1) \times G_2) \cup (G_1 \times (G_2 - E_2)) . \tag{15.49}$$

By Proposition 15.9, we conclude that $m^* (G - E) = 0$, hence E is measurable. ■

Proposition 15.11

Suppose f is a measurable function on $\mathbb{R}^{d_1}$. Then the function $\widetilde{f}$ defined by $\widetilde{f} (x, y) = f (x)$ is measurable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

Proof. We assume f is real-valued. If $a \in \mathbb{R}$ and $E_1 = \left\{ x \in \mathbb{R}^{d_1} : f (x) < a \right\}$, then E_1 is measurable by definition. Since

$$\left\{ (x, y) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} : \widetilde{f} (x, y) < a \right\} = E_1 \times \mathbb{R}^{d_2}, \tag{15.50}$$

which is measurable by Proposition 15.10. So $\widetilde{f}^{-1}(-\infty, a)$ is measurable for every $a \in \mathbb{R}$. Thus $\widetilde{f}$ is a measurable function on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, as desired. ■

Proposition 15.12

Suppose f is a non-negative function in $\mathbb{R}^d$ and let

$$\mathcal{A} = \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : 0 \leq y \leq f(x)\right\}. \tag{15.51}$$

Then

(i) f is measurable in $\mathbb{R}^d$ if and only if $\mathcal{A}$ is a measurable set on $\mathbb{R}^{d+1}$.

(ii) If the conditions in (i) hold then

$$\int_{\mathbb{R}^d} f(x)\, \mathrm{d}x = m(\mathcal{A}). \tag{15.52}$$

Proof. (i) If f is measurable on $\mathbb{R}^d$, then Proposition 15.11 guarantees that the function

$$F(x, y) = y - f(x) \tag{15.53}$$

is measurable on $\mathbb{R}^{d+1}$, so we immediately see that $\mathcal{A} = \{y \geq 0\} \cap \{F \leq 0\}$ is measurable.

Conversely, suppose that $\mathcal{A}$ is measurable. We note that for each $x \in \mathbb{R}^{d_1}$ the slice $\mathcal{A}_x = \{y \in \mathbb{R} : (x, y) \in \mathcal{A}\}$ is a closed segment, namely $\mathcal{A}_x = [0, f(x)]$. Consequently, Corollary 15.8 (with the roles of x and y interchanged) yields the measurability of

$$x \mapsto m(\mathcal{A}_x) = f(x). \tag{15.54}$$

(ii) By Corollary 15.8 (again with interchanging x and y),

$$m(\mathcal{A}) = \int_{\mathbb{R}} m(\mathcal{A}_x)\, \mathrm{d}x = \int_{\mathbb{R}} f(x)\, \mathrm{d}x. \tag{15.55}$$

■

Proposition 15.13

If $E \subseteq \mathbb{R}^d$ is a measurable set then the set

$$\widetilde{E} = \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : x - y \in E\right\} \tag{15.56}$$

is measurable in $\mathbb{R}^d \times \mathbb{R}^d$.

Proof. Note first that if O is an open set, then $\widetilde{O}$ is also open. Taking countable intersections shows that if E is a G_δ set, then so is $\widetilde{E}$, because

$$\widetilde{\bigcap_{n=1}^{\infty} O_n} = \bigcap_{n=1}^{\infty} \widetilde{O_n}, \tag{15.57}$$

where $O_n \subseteq \mathbb{R}^d$ open. Now, if $m(E) = 0$, we claim that $\widetilde{E}$ is measurable with $m\left(\widetilde{E}\right) = 0$. Take a G_δ-set G containing E such that $m(G - E) = 0$. Since $m(E) = 0$, we have $m(G) = 0$.

By Tonelli's theorem,

$$m\left(\widetilde{G}\right) = \int_{\mathbb{R}^{2d}} \chi_{\widetilde{G}}(x, y) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \chi^y_{\widetilde{G}}(x)\, \mathrm{d}x \right) \mathrm{d}y. \tag{15.58}$$

Now, $\chi^y_{\widetilde{G}}(x) = 1$ if and only if $(x, y) \in \widetilde{G}$ if and only if $x - y \in G$. Therefore,

$$m\left(\widetilde{E}\right) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \chi_G(x - y)\, \mathrm{d}x \right) \mathrm{d}y. \tag{15.59}$$

By translation-invariance of Lebesgue measure,

$$\int_{\mathbb{R}^d} \chi_G(x - y)\, \mathrm{d}x = \int_{\mathbb{R}^d} \chi_G(u)\, \mathrm{d}u = m(G) = 0. \tag{15.60}$$

As a result, $m\left(\widetilde{G}\right) = 0$. Since Lebesgue measure is complete, $\widetilde{E} \subseteq \widetilde{G}$ is also measurable, and it has measure 0.

Now, for a general measurable set E, there exists a G_δ-set $G \supseteq E$ such that $m(G - E) = 0$. Now, $\widetilde{G}$ is a G_δ-set containing $\widetilde{E}$, and

$$m\left(\widetilde{G} - \widetilde{E}\right) = m\left(\widetilde{G - E}\right) = 0, \tag{15.61}$$

since $m(G - E) = 0$. Hence, $\widetilde{E}$ is measurable! ∎

Proposition 15.14

If f is a measurable function on $\mathbb{R}^d$ the function

$$\widetilde{f}(x, y) = f(x - y) \tag{15.62}$$

is measurable in $\mathbb{R}^d \times \mathbb{R}^d$.

Proof. Let $E = \left\{z \in \mathbb{R}^d : f(z) < a\right\}$, which is measurable since f is a measurable function on $\mathbb{R}^d$. Now,

$$\left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \widetilde{f}(x, y) < a\right\} = \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : f(x - y) < a\right\} \tag{15.63}$$

$$= \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : x - y \in E\right\} = \widetilde{E}, \tag{15.64}$$

which is measurable by Proposition 15.13. So $\widetilde{f}^{-1}(-\infty, a)$ is measurable for every $a \in \mathbb{R}$. Thus $\widetilde{f}$ is a measurable function. ■

§15.iii. The case of general measures

Now let $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ be general measure spaces. Let us assume that each of them is σ-finite. We want to describe what the "product σ-algebra" and the "product measure" $\mu = \mu_1 \times \mu_2$ are in the product space

$$\Omega = \Omega_1 \times \Omega_2 = \{(x_1, x_2) : x_1 \in \Omega_1,\, x_2 \in \Omega_2\}\,. \tag{15.65}$$

We start by defining **measurable rectangles**. These are subsets of Ω of the form $A \times B$ where $A \in \Sigma_1$ and $B \in \Sigma_2$. Let $\mathcal{A}$ be the collection of all sets in Ω that are finite unions of disjoint measurable rectangles. One can check that $\mathcal{A}$ is an algebra of subsets of Ω. The following result will prove that $\mathcal{A}$ is an algebra, since for $A \subseteq \Omega_1$, $B \subseteq \Omega_2$,

$$(A \times B)^c = (\Omega_1 \times B^c) \cup (A^c \times B)\,; \tag{15.66}$$

and for $A, C \subseteq \Omega_1$, $B, D \subseteq \Omega_2$,

$$(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)\,. \tag{15.67}$$

Proposition 15.15

We define an **elementary family** to be a collection $\mathcal{E}$ of subsets of Ω such that

1. $\varnothing \in \mathcal{E}$,
2. if $E, F \in \mathcal{E}$ then $E \cap F \in \mathcal{E}$,
3. if $E \in \mathcal{E}$ then E^c is a finite disjoint union of members of $\mathcal{E}$.

If $\mathcal{E}$ is an elementary family, the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.

Proof. If $A, B \in \mathcal{E}$ and $B^c = \bigcup_{j=1}^n C_j$, where $C_j \in \mathcal{E}$, disjoint, then

$$A - B = \bigcup_1^n (A \cap C_j) \tag{15.68}$$

and

$$A \cup B = (A - B) \cup B, \tag{15.69}$$

where these unions are disjoint, so $A - B, A \cup B \in \mathcal{A}$. It now follows by induction that if

$A_1, \ldots, A_n \in \mathcal{E}$, then

$$\bigcup_{j=1}^{n} A_j \in \mathcal{A}. \tag{15.70}$$

Indeed, by inductive hypothesis we may assume that $A_1, \ldots, A_{n-1}$ are disjoint, and then

$$\bigcup_{j=1}^{n} A_j = A_n \cup \bigcup_{j=1}^{n-1} (A_j - A_n) \tag{15.71}$$

which is a disjoint union. To see that $\mathcal{A}$ is closed under complements, suppose $A_1, \ldots, A_n \in \mathcal{E}$ and $A_m^c = \bigcup_{j=1}^{N_m} B_m^j$ with $B_m^1, \ldots, B_m^{N_m}$ disjoint members of $\mathcal{E}$. Then

$$\left(\bigcup_{m=1}^{n} A_m\right)^c = \bigcap_{m=1}^{n} \left(\bigcup_{j=1}^{N_m} B_m^j\right) = \bigcup \left\{ B_1^{j_1} \cap \cdots \cap B_n^{j_n} : 1 \leq j_m \leq N_m,\, 1 \leq m \leq n \right\}, \tag{15.72}$$

which is in $\mathcal{A}$. ■

From now on we abbreviate our terminology and call the measurable rectangles simply *rectangles*.

On rectangles we define the function

$$\mu_0 (A \times B) = \mu_1 (A)\, \mu_2 (B). \tag{15.73}$$

In fact, the function μ_0 has a unique extension to the algebra $\mathcal{A}$ and it becomes a pre-measure on $\mathcal{A}$. This is a consequence of the following fact:

Lemma 15.16

Whenever a rectangle $A \times B$ is a disjoint union of countably many rectangles $A_j \times B_j$ we have

$$\mu_0 (A \times B) = \sum_{j=1}^{\infty} \mu_0 (A_j \times B_j). \tag{15.74}$$

Proof. Fix $x_1 \in A$. Then, for every $x_2 \in B$ we have that (x_1, x_2) belongs to exactly one $A_j \times B_j$. Then B is the disjoint union of the B_j's such that $x_1 \in A_j$ (call this subset of indices $J(x_1)$). By the countable additivity of the measure μ_2 we have

$$\chi_A (x_1)\, \mu_2 (B) = \sum_{j \in J(x_1)} \mu_2 (B_j) = \sum_{j=1}^{\infty} \chi_{A_j} (x_1)\, \mu_2 (B_j). \tag{15.75}$$

We now integrate in x_1 and use Proposition 2.8 (which is a direct consequence of Monotone Convergence Theorem) to get

$$\mu_1 (A)\, \mu_2 (B) = \sum_{j=1}^{\infty} \mu_1 (A_j)\, \mu_2 (B_j). \tag{15.76}$$

■

Using Theorem 7.4 we find that there exists a measure that we call $\mu = \mu_1 \times \mu_2$ on the σ-algebra Σ generated by $\mathcal{A}$ that extends μ_0. Since we are assuming that μ_1 and μ_2 are σ-finite, so will μ_0 be, and hence such an extension μ to Σ is in fact unique. This defines our product measure space $(\Omega, \Sigma, \mu) = (\Omega_1 \times \Omega_2, \Sigma, \mu_1 \times \mu_2)$.

Given now $E \in \Sigma$ we consider the slices

$$E_{x_1} = \{x_2 \in \Omega_2 : (x_1, x_2) \in E\} \qquad \text{and} \qquad E^{x_2} = \{x_1 \in \Omega_1 : (x_1, x_2) \in E\}. \tag{15.77}$$

Let $\mathcal{A}_\sigma$ be the collection of sets that are countable unions of sets in $\mathcal{A}$, and $\mathcal{A}_{\sigma\delta}$ the collection of sets that are countable intersections of sets in $\mathcal{A}_\sigma$. We now prove the following key fact.

Proposition 15.17

If E is a measurable set in the product σ-algebra Σ, then E^{x_2} is μ_1-measurable for μ_2-almost every $x_2 \in \Omega_2$. Moreover, $\mu_1(E^{x_2})$ is a μ_2-measurable function and

$$\int_{\Omega_2} \mu_1(E^{x_2})\, \mathrm{d}\mu_2(x_2) = (\mu_1 \times \mu_2)(E). \tag{15.78}$$

Proof. **Step 1.** We first prove that the result holds for every $x_2 \in \Omega_2$ when E belongs to $\mathcal{A}_{\sigma\delta}$. Observe first that the assertion holds if E is a measurable rectangle. Suppose that E is a set in $\mathcal{A}_\sigma$. We can write E as a countable disjoint union of rectangles E_j. Then

$$E^{x_2} = \bigcup_{j=1}^{\infty} E_j^{x_2} \tag{15.79}$$

and we observe that $\left\{E_j^{x_2}\right\}$ are disjoint sets. Then we may apply (15.78) to each E_j and sum up using Monotone Convergence Theorem to get the result for E.

Now let us consider the case $E \in \mathcal{A}_{\sigma\delta}$. Since we are assuming that μ_1 and μ_2 are σ-finite, there are sequences $F_1 \subseteq F_2 \subseteq F_3 \subseteq \ldots \subseteq \Omega_1$ and $G_1 \subseteq G_2 \subseteq G_3 \subseteq \ldots \subseteq \Omega_2$ such that $\bigcup_j F_j = \Omega_1$, $\bigcup_j G_j = \Omega_2$, $\mu_1(F_j) < \infty$, $\mu_2(G_j) < \infty$ for all j. We work with the set

$$A_j = E \cap (F_j \times G_j). \tag{15.80}$$

If we can prove the result for such A_j then the general result will follow simply by the Monotone Convergence Theorem and the fact that A_j increases to E.

Therefore, assume that our set $E \in \mathcal{A}_{\sigma\delta}$ is such that $E \subseteq F \times G$ with $\mu_1(F) < \infty$, $\mu_2(G) < \infty$. Then there is a sequence of sets $\{E_j\} \subseteq \mathcal{A}_\sigma$ with $E_j \supset E_{j+1}$ such that

$$E = \bigcap_{j=1}^{\infty} E_j. \tag{15.81}$$

By intersecting the sequence with the rectangle $F \times G$, we may assume that $E_1 \subseteq F \times G$. Let $f_j(x_2) = \mu_1\left(E_j^{x_2}\right)$ and $f(x_2) = \mu_1\left(E^{x_2}\right)$. Note that

$$E^{x_2} = \bigcap_{j=1}^{\infty} E_j^{x_2} \tag{15.82}$$

and hence E^{x_2} is μ_1-measurable. Since $(\mu_1 \times \mu_2)(E_1) < \infty$ (that is why we asked $E_1 \subseteq F \times G$), we see that $f_j \to f$ as $j \to \infty$ for each x_2 and hence f is μ_2-measurable. Since $\{f_j\}$ is a decreasing sequence of non-negative functions we have

$$\int_{\Omega_2} f(x_2)\, \mathrm{d}\mu_2 = \lim_{j\to\infty} \int_{\Omega_2} f_j(x_2)\, \mathrm{d}\mu_2 \tag{15.83}$$

by the dominated convergence theorem, since f_1 is integrable. This establishes (15.78) in this case and we are done.

Step 2. We now consider the general case. Recall that the product measure μ arises as an outer measure μ^* generated by the algebra $\mathcal{A}$. In particular, if E is measurable, for any $\varepsilon > 0$ there is a union $E \subseteq B_\varepsilon = \bigcup_j A_j$, with $A_j \in \mathcal{A}$, that we may take disjoint, such that

$$\mu(B_\varepsilon) = \sum_{j=1}^{\infty} \mu(A_j) \leq \mu(E) + \varepsilon. \tag{15.84}$$

By taking $\varepsilon = \frac{1}{n}$ and letting $C = \bigcap_n B_{1/n}$, we find that $C \in \mathcal{A}_{\sigma\delta}$, $E \subseteq C$ and

$$\mu(C) = \mu(E). \tag{15.85}$$

As in the final part of the proof for Step 1, we can assume that our set E is such that $\mu(E) < \infty$. This leads us to a representation $C = E \cup Z$ (disjoint union) with $\mu(Z) = 0$. Here $Z = C - E$.

Let us first prove the result for a set Z such that $\mu(Z) = 0$. As we argued above, there exists a set $\tilde{Z} \in \mathcal{A}_{\sigma\delta}$ such that $Z \subseteq \tilde{Z}$ and $\mu\left(\tilde{Z}\right) = 0$. Since $Z^{x_2} \subseteq \tilde{Z}^{x_2}$ for every x_2 and, by (15.78), the function $\mu_1\left(\tilde{Z}^{x_2}\right)$ is zero for μ_2-almost every x_2, the assumed completeness of the measure μ_1 guarantees that Z^{x_2} is μ_1-measurable and $\mu_1\left(Z^{x_2}\right) = 0$ for such μ_2-almost every x_2. The desired conclusion then holds for Z.

Coming back to the case of measurable E with $\mu(E) < \infty$, we wrote $C = E \cup Z$ (disjoint union) with $C \in \mathcal{A}_{\sigma\delta}$ and $\mu(Z) = 0$. Then $C^{x_2} = E^{x_2} \cup Z^{x_2}$, which implies that $E^{x_2} = C^{x_2} - Z^{x_2}$. Hence E^{x_2} is μ_1-measurable for μ_2-almost every x_2 and $\mu_1\left(E^{x_2}\right) = \mu_1\left(C^{x_2}\right)$, from which the proposition follows. ■

Theorem 15.18 (Fubini's theorem for general measuere spaces)

Assume that $f(x_1, x_2)$ is an integrable function in $(\Omega_1 \times \Omega_2, \mu_1 \times \mu_2)$. Then

(i) For μ_2-almost every x_2, the slice $f^{x_2}(x_1) := f(x_1, x_2)$ is integrable in (Ω_1, μ_1).

(ii) The function $x_2 \mapsto \int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\mu_1(x_1)$ is integrable in (Ω_2, μ_2).

(iii) We have

$$\int_{\Omega_2} \left(\int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) = \int_{\Omega_1 \times \Omega_2} f\, \mathrm{d}(\mu_1 \times \mu_2). \tag{15.86}$$

Proof. If the desired conclusions hold for finitely many functions, they hold for their linear combinations. We may hence assume that f is non-negative. If f is the characteristic function of a measurable set E, that is $f = \chi_E$, we have already proved the result in Proposition 15.17. Then we have the result for simple functions, and by Monotone Convergence Theorem for all non-negative integrable functions. ■

We have a similar statement when all functions involved are non-negative.

Theorem 15.19 (Tonelli's theorem for general measuere spaces)

Assume that $f(x_1, x_2)$ is a non-negative and measurable function in $(\Omega_1 \times \Omega_2, \mu_1 \times \mu_2)$. Then

(i) For μ_2-almost every x_2, the slice $f^{x_2}(x_1) := f(x_1, x_2)$ is measurable in (Ω_1, μ_1).

(ii) The function $x_2 \mapsto \int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\mu_1(x_1)$ is measurable in (Ω_2, μ_2).

(iii) We have

$$\int_{\Omega_2} \left(\int_{\Omega_1} f(x_1, x_2)\, \mathrm{d}\mu_1(x_1) \right) \mathrm{d}\mu_2(x_2) = \int_{\Omega_1 \times \Omega_2} f\, \mathrm{d}(\mu_1 \times \mu_2). \tag{15.87}$$

Proof. Just follow the argument in the proof of Tonelli's theorem using that the measures are σ-finite. ■

Remark 15.6. We conclude by remarking that the product space (Ω, Σ, μ) that we constructed above is not necessarily complete. If we consider the completed space $\left(\Omega, \overline{\Sigma}, \overline{\mu}\right)$ in the sense of Remark 1.14, Fubini's theorem continues to hold. The proof requires a minor adjustment in Step 2 of the proof of Proposition 15.17.

§15.iv. Exercises

Exercise 15.1. Let $L : \mathbb{R}^d \to \mathbb{R}^d$ be a linear transformation. If $E \subseteq \mathbb{R}^d$ is a Lebesgue measurable set then show that $L(E)$ is also Lebesgue measurable, and

$$m(L(E)) = |\det L|\, m(E). \tag{15.88}$$

Solution. Although we have solved this problem in Theorem 6.11, here's another way of solving it. This time, since we have more tools in our toolbox (e.g. Fubini's theorem and Tonelli's theorem[2]), we will make use of those.

If E is a compact set, then so is $L(E)$ (continuous image of compact set is compact). Since a F_σ-set is a countable union of closed sets, and every closed set can be written as a countable union of compact sets (namely $C \cap \overline{B(0,n)}$ is compact for C closed), therefore, every F_σ set is countable union of compact sets. Hence, if $F = \bigcup_{n=1}^{\infty} K_n$ is a F_σ set with K_n being compact,

$$L(F) = L\left(\bigcup_{n=1}^{\infty} K_n\right) = \bigcup_{n=1}^{\infty} L(K_n) \tag{15.89}$$

is an F_σ set.

Now, suppose $m(E) = 0$. Since L is a linear map between finite dimensional vector spaces, it's bounded, i.e.

$$\left|L(x) - L\left(x'\right)\right| < M\left|x - x'\right|, \tag{15.90}$$

for some constant M. Suppose Q is a cube with side-length l. Then for $x, x' \in Q$, $|x - x'| \leq \sqrt{d}l$, so that

$$\left|L(x) - L\left(x'\right)\right| < \sqrt{d}Ml. \tag{15.91}$$

Therefore, $L(Q) \subseteq Q'$, where Q' is the cube centered $L(x)$ with side length $2\sqrt{d}Ml$.

$$\left|Q'\right| = c\left|Q\right|, \tag{15.92}$$

where $c = \left(2\sqrt{d}M\right)^c$. Therefore, since $m(E) = 0$, we can find cubes Q_i such that $E \subseteq \bigcup_{i=1}^{\infty} Q_i$ and

$$\sum_{i=1}^{\infty} |Q_i| \leq \varepsilon. \tag{15.93}$$

Now, $L(E) \subseteq \bigcup_{i=1}^{\infty} L(Q_i) \subseteq \bigcup_{i=1}^{\infty} Q_i'$, so

$$m^*(L(E)) \leq \sum_{i=1}^{\infty} \left|Q_i'\right| = c\sum_{i=1}^{\infty} |Q_i| = c\varepsilon. \tag{15.94}$$

Therefore, $m^*(L(E)) = 0$, if $m(E) = 0$.

Now, if E is any measurable set, $E = F \cup N$ for some F_σ set F and null set N. Then

$$L(E) = L(F \cup N) = L(F) \cup L(N), \tag{15.95}$$

and we have shown that $L(F)$ is an F_σ set, and $L(N)$ is a null set. Hence, $L(E)$ is measurable.

Now, we'll show that $m(L(E)) = |\det L|\, m(E)$. Let's first consider the case $\det L = 0$. As before (Proposition 6.12), we'll show that if $H = L\left(\mathbb{R}^d\right)$, $m(H) = 0$.

[2] As Abraham Kaplan said, "Give a small boy a hammer, and he will find that everything he encounters needs pounding."

Since H is a stricly lower-dimensional hyperplane through the origin, there exists some nonzero vector $\mathbf{v} \in \mathbb{R}^d$ such that $\mathbf{x} \cdot \mathbf{v} = 0$ for every $\mathbf{x} \in H$. WLOG, we'll assume $v_d \neq 0$. Then for $\mathbf{x} \in H$,

$$\sum_{i=1}^{d} x_i v_i = 0 \implies x_d = -\frac{1}{v_d} \sum_{i=1}^{d-1} x_i v_i. \tag{15.96}$$

Therefore, we have

$$\chi_H(x_1, \ldots, x_d) \leq \chi_{\left\{x_d = -\frac{1}{v_d} \sum_{i=1}^{d-1} x_i v_i\right\}}(x_1, \ldots, x_d). \tag{15.97}$$

Let $\mathrm{d}\mathbf{x}$ be the d-dimensional Lebesgue measure, and $\mathrm{d}\mathbf{x}^{d-1}$ is the $(d-1)$-dimensional Lebesgue measure. Then by Tonelli's theorem,

$$m(H) \leq \int_{\mathbb{R}^d} \chi_{\left\{x_d = -\frac{1}{v_d} \sum_{i=1}^{d-1} x_i v_i\right\}} \mathrm{d}\mathbf{x} = \int_{\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \chi_{\left\{x_d = -\frac{1}{v_d} \sum_{i=1}^{d-1} x_i v_i\right\}} \mathrm{d}x_d \right) \mathrm{d}\mathbf{x}^{d-1}. \tag{15.98}$$

With everything else being fixed, the inner integral is actually $\int_{\mathbb{R}} \chi_{\text{singleton}} \, \mathrm{d}x_d$, which is clearly 0 since singletons have 0 Lebesgue measure. Therefore,

$$m\left(L\left(\mathbb{R}^d\right)\right) = 0 \text{ when } \det L = 0. \tag{15.99}$$

Now suppose $\det L \neq 0$. As before, we'll show (15.88) for "elementary" matrices of the following types:

1. Permutation type: $L_1(\ldots, x_i, x_{i+1}, \ldots) = (\ldots, x_{i+1}, x_i, \ldots)$;
2. Scaling type: $L_2(x_1, \ldots, x_d) = (cx_1, x_2, \ldots, x_d)$ for some nonzero c;
3. Shearing type: $L_3(x_1, \ldots, x_d) = (x_1 + x_2, x_2, \ldots, x_d)$ for some nonzero c;

Case 1: Permutation type: In this case, L swaps x_i and x_{i+1}, so we have

$$\chi_{L(E)}(\ldots, x_i, x_{i+1}, \ldots) = \chi_E(\ldots, x_{i+1}, x_i, \ldots) \tag{15.100}$$

Hence, by Tonelli's theorem,

$$m(L(E)) = \int_{\mathbb{R}^d} \chi_{L(E)} = \int_{\mathbb{R}^{d-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_E(\ldots, x_{i+1}, x_i, \ldots) \, \mathrm{d}x_i \, \mathrm{d}x_{i+1} \, \mathrm{d}\mathbf{x}^{d-2}. \tag{15.101}$$

Let $x_{i+1} = y_i$ and $x_i = y_{i+1}$, (and swap the order of integration) so that

$$m(L(E)) = \int_{\mathbb{R}^{d-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \chi_E(\ldots, y_i, y_{i+1}, \ldots) \, \mathrm{d}y_i \, \mathrm{d}y_{i+1} \, \mathrm{d}\mathbf{x}^{d-2} = \int_{\mathbb{R}^d} \chi_E = m(E). \tag{15.102}$$

Therefore, (15.88) holds for permutation type linear maps.

Case 2: Scaling type: In this case, L scales the first coordinate, so we have

$$\chi_{L(E)}(x_1, \ldots, x_d) = \chi_E\left(\frac{x_1}{c}, \ldots, x_d\right). \tag{15.103}$$

Hence, by Tonelli's theorem,

$$m(L(E)) = \int_{\mathbb{R}^d} \chi_{L(E)} = \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \chi_{(E)}\left(\frac{x_1}{c}, \ldots, x_d\right) \mathrm{d}x_1 \, \mathrm{d}\mathbf{x}^{d-1}. \tag{15.104}$$

Let's do the 1-dimensional change of variables $y_1 = \frac{x_1}{c}$, so that $|c| \, \mathrm{d}y_1 = \mathrm{d}x_1$. Therefore,

$$m(L(E)) = |c| \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \chi_{(E)}(y_1, \ldots, x_d) \, \mathrm{d}y_1 \, \mathrm{d}\mathbf{x}^{d-1} = |c| \int_{\mathbb{R}^d} \chi_E = |c| \, m(E). \tag{15.105}$$

Therefore, (15.88) holds for scaling type linear maps.

Case 3: Shearing type: In this case, L adds x_2 to x_1, so we have

$$\chi_{L(E)}(x_1, x_2, \ldots, x_d) = \chi_E(x_1 - x_2, \ldots, x_d). \tag{15.106}$$

Hence, by Tonelli's theorem,

$$m(L(E)) = \int_{\mathbb{R}^d} \chi_{L(E)} = \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \chi_E(x_1 - x_2, \ldots, x_d) \, \mathrm{d}x_1 \, \mathrm{d}\mathbf{x}^{d-1}. \tag{15.107}$$

For a fixed x_2, we do the 1-dimensional change of variables $y_1 = x_1 - x_2$ so that $\mathrm{d}x_1 = \mathrm{d}y_1$.

$$m(L(E)) = \int_{\mathbb{R}^{d-1}} \int_{\mathbb{R}} \chi_E(y_1, \ldots, x_d) \, \mathrm{d}y_1 \, \mathrm{d}\mathbf{x}^{d-1} = \int_{\mathbb{R}^d} \chi_E = |c| \, m(E). \tag{15.108}$$

Therefore, (15.88) holds for shearing type linear maps.

Now, if (15.88) holds for L_1 and L_2, then for any measurable set E,

$$m(L_1 L_2(E)) = |\det L_1| \, m(L_2 E) = |\det L_1| \, |\det L_2| \, m(E) = |\det(L_1 L_2)| \, m(E); \tag{15.109}$$

so it holds for $L_1 L_2$ as well! Since the elementary matrices generate the whole $\mathrm{GL}(d, \mathbb{R})$, we can conclude that (15.88) holds for all invertible linear maps! ■

Exercise 15.2 (Minkowski's inequality for integrals, aka. triangle inequality)**.** Let $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ be σ-finite measure spaces, and let f be a measurable function in the product space $\Omega_1 \times \Omega_2$. For $1 \leq p \leq \infty$ show that

$$\left(\int_{\Omega_1} \left(\int_{\Omega_2} |f(x,y)| \, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \leq \int_{\Omega_2} \left(\int_{\Omega_1} |f(x,y)|^p \, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y). \tag{15.110}$$

Solution. $|f(x,y)|$ is a non-negative function, so for $p = 1$, the statement is just Tonelli's theorem for general measure spaces. For $p = \infty$, the statement is

$$\operatorname*{ess\,sup}_{x \in \Omega_1} \int_{\Omega_2} |f(x, -)| \, \mathrm{d}\mu_2 \leq \int_{\Omega_2} \operatorname*{ess\,sup}_{x \in \Omega_1} |f(x, -)| \, \mathrm{d}\mu_2 \tag{15.111}$$

By definition of essential supremum, for each fixed $y \in \Omega_2$,

$$|f(x,y)| \leq \operatorname*{ess\,sup}_{x\in\Omega_1} |f(x,y)| \tag{15.112}$$

for μ_1-almost every x. Integrating with respect to μ_2, we get

$$\int_{\Omega_2} |f(x,y)| \, \mathrm{d}\mu_2(y) \leq \int_{\Omega_2} \operatorname*{ess\,sup}_{x\in\Omega_1} |f(x,y)| \, \mathrm{d}\mu_2(y). \tag{15.113}$$

The RHS is a constant independent of x, which is greater than (or equal to) the LHS for μ_1-almost every x. Hence, it's greater than (or equal to) the essential supremum of the LHS:

$$\operatorname*{ess\,sup}_{x\in\Omega_1} \int_{\Omega_2} |f(x,y)| \, \mathrm{d}\mu_2(y) \leq \int_{\Omega_2} \operatorname*{ess\,sup}_{x\in\Omega_1} |f(x,y)| \, \mathrm{d}\mu_2(y). \tag{15.114}$$

Now let's assume $1 < p < \infty$. Let I be the LHS of (15.110). If $I = 0$, then we have nothing to prove. For now, assume $I < \infty$ (we'll deal with the case $I = \infty$ case later). Let

$$g(x) := \int_{\Omega_2} |f(x,-)| \, \mathrm{d}\mu_2. \tag{15.115}$$

Then

$$I^p = \int_{\Omega_1} g^p \, \mathrm{d}\mu_1 = \int_{\Omega_1} g^{p-1} \left(\int_{\Omega_2} |f(x,-)| \, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 \tag{15.116}$$

$$= \int_{\Omega_2} \left(\int_{\Omega_1} g^{p-1} \, |f(x,-)| \, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2, \tag{15.117}$$

where the switch of integrals is fascilitated by Tonelli's theorem for general measuere spaces. Now we use Hölder's inequality in the inner integral with respect to p and q for $\frac{1}{p} + \frac{1}{q} = 1$.

$$\int_{\Omega_1} g^{p-1} \, |f(x,-)| \, \mathrm{d}\mu_1 \leq \left(\int_{\Omega_1} |f(x,-)|^p \, \mathrm{d}\mu_1 \right)^{\frac{1}{p}} \left(\int_{\Omega_1} |g|^{(p-1)q} \, \mathrm{d}\mu_1 \right)^{\frac{1}{q}} \tag{15.118}$$

Since $(p-1)q = p$ and $\frac{1}{q} = 1 - \frac{1}{p} = \frac{p-1}{p}$,

$$\left(\int_{\Omega_1} |g|^{(p-1)q} \, \mathrm{d}\mu_1 \right)^{\frac{1}{q}} = \left(\int_{\Omega_1} |g|^p \, \mathrm{d}\mu_1 \right)^{\frac{p-1}{p}} = I^{p-1}. \tag{15.119}$$

Plugging things into (15.117),

$$I^p \leq \int_{\Omega_2} \left[\left(\int_{\Omega_1} |f(x,-)|^p \, \mathrm{d}\mu_1 \right)^{\frac{1}{p}} I^{p-1} \right] \mathrm{d}\mu_2 = I^{p-1} \int_{\Omega_2} \left(\int_{\Omega_1} |f(x,-)|^p \, \mathrm{d}\mu_1 \right)^{\frac{1}{p}} \mathrm{d}\mu_2. \tag{15.120}$$

Cancelling out I^{p-1} (which is finite, by assumption),

$$I \leq \int_{\Omega_2} \left(\int_{\Omega_1} |f(x,-)|^p \, \mathrm{d}\mu_1 \right)^{\frac{1}{p}} \mathrm{d}\mu_2, \tag{15.121}$$

as desired.

Now suppose $I = \infty$. Since the measure spaces are σ-finite, we can find an increasing sequence of finite-measure measurable sets

$$E_1 \subseteq E_2 \subseteq \cdots \subseteq \Omega_1, \text{ and } F_1 \subseteq F_2 \subseteq \cdots \subseteq \Omega_2, \tag{15.122}$$

such that

$$\Omega_1 = \bigcup_{n=1}^{\infty} E_n, \text{ and } \Omega_2 = \bigcup_{n=1}^{\infty} F_n. \tag{15.123}$$

For each n, define

$$f_n(x,y) = \min\{n, |f(x,y)|\}\, \chi_{E_n}(x)\chi_{F_n}(y). \tag{15.124}$$

Then $f_n \to |f|$ pointwise everywhere, and $f_n \leq f_{n+1}$. Since (15.110) holds for LHS being finite, we have

$$\left(\int_{\Omega_1}\left(\int_{\Omega_2} |f_n(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1} |f_n(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y), \tag{15.125}$$

since its LHS is at most

$$\left[\left(n \cdot \mu_2(F_n)\right)^p \mu_1(E_n)\right]^{\frac{1}{p}}. \tag{15.126}$$

Now we apply Monotone Convergence Theorem on both sides of (15.125) to pass the limit inside the integral to get

$$\left(\int_{\Omega_1}\left(\int_{\Omega_2} |f(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \leq \int_{\Omega_2}\left(\int_{\Omega_1} |f(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y). \tag{15.127}$$

To see the limit passing inside in action, for LHS,

$$\begin{aligned}
\lim_{n\to\infty}\left(\int_{\Omega_1}\left(\int_{\Omega_2} |f_n(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} &= \left(\int_{\Omega_1}\lim_{n\to\infty}\left(\int_{\Omega_2} |f_n(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \\
&= \left(\int_{\Omega_1}\left(\int_{\Omega_2}\lim_{n\to\infty} |f_n(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \\
&= \left(\int_{\Omega_1}\left(\int_{\Omega_2} |f(x,y)|\, \mathrm{d}\mu_2(y)\right)^p \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}};
\end{aligned}$$

and for RHS,

$$\begin{aligned}
\lim_{n\to\infty}\int_{\Omega_2}\left(\int_{\Omega_1} |f_n(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y) &= \int_{\Omega_2}\lim_{n\to\infty}\left(\int_{\Omega_1} |f_n(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1}\lim_{n\to\infty} |f_n(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y) \\
&= \int_{\Omega_2}\left(\int_{\Omega_1} |f(x,y)|^p\, \mathrm{d}\mu_1(x)\right)^{\frac{1}{p}} \mathrm{d}\mu_2(y),
\end{aligned}$$

since raising p-th power and $\frac{1}{p}$-th power preserves monotonicity. So we are done! ■

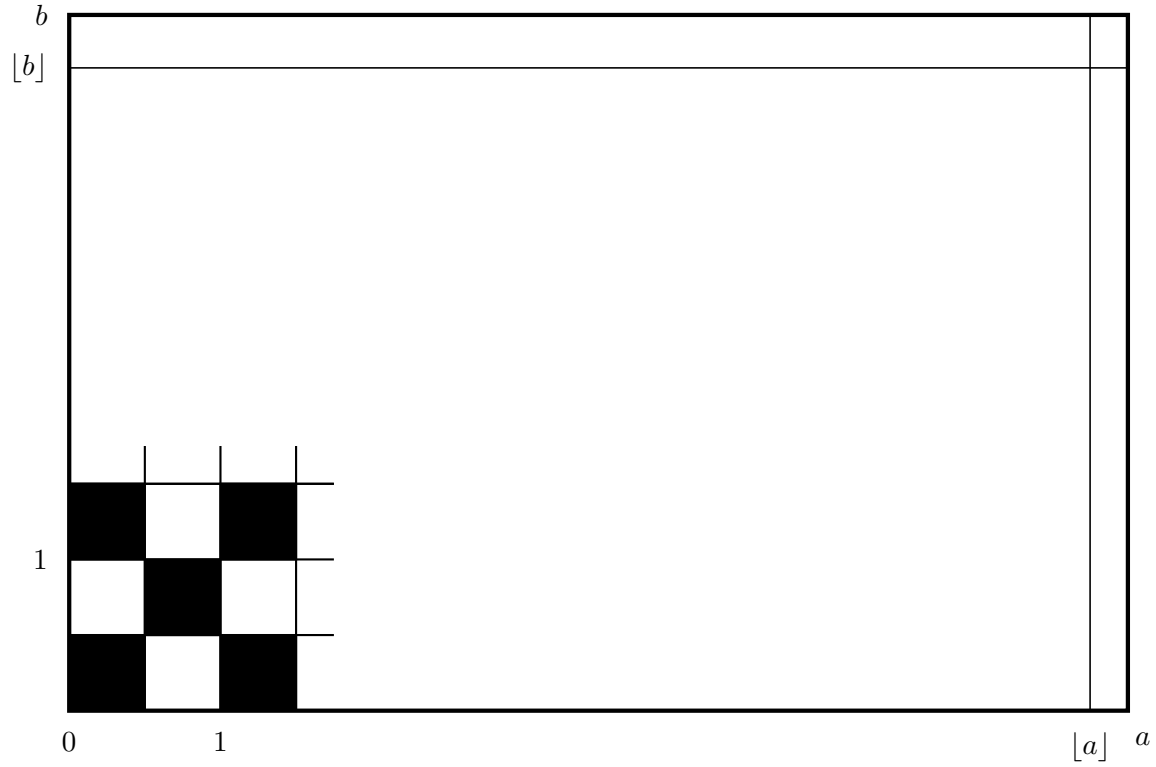
Exercise 15.3. Let R be a (say, closed) rectangle in the plane. Assume that R is the almost disjoint union of a finite number of smaller rectangles $\{R_j\}_{j=1}^N$, (R and all R_j's with sides parallel to the coordinate axes). Assume that each rectangle R_j has at least one side with an integer length. Prove or disprove: R has one side with an integer length.

Solution. We divide the plane into squares of size $\frac{1}{2} \times \frac{1}{2}$ and color them in a chess-board pattern.

We claim that

> A rectangle R has equal amount of black and white area if and only if it has at least one side-length integer.

WLOG, assume one corner of R is the origin (if not, we just translate so that it happens), and R lies in the first quadrant ($x, y \geq 0$). Let $R = [0, a] \times [0, b]$. It's clear that if R has an integer side, then it contains an equal amount of black and white area. Conversely, suppose R doesn't have any integer-length side. Then the picture looks something like this:



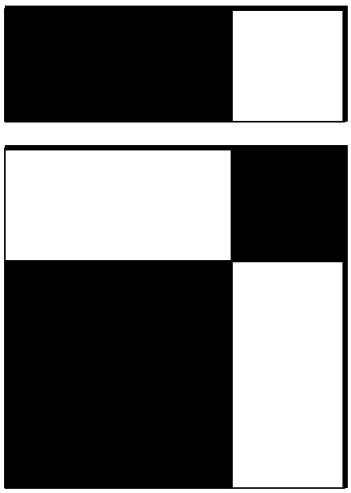
If R has no integer length, then in the above picture, the sub-rectangles

$$[0, \lfloor a \rfloor] \times [0, \lfloor b \rfloor], \quad [\lfloor a \rfloor, a] \times [0, \lfloor b \rfloor], \quad [0, \lfloor a \rfloor] \times [\lfloor b \rfloor, b], \tag{15.128}$$

have the same amount of black and white area (since they have at least one side-length integer). Therefore, the top-right subrectangle $R' := [\lfloor a \rfloor, a] \times [\lfloor b \rfloor, b]$ also has an equal amount of black and white area.

If $a - \lfloor a \rfloor, b - \lfloor b \rfloor \leq \frac{1}{2}$, then R' is entirely contained within a grid-square, which is colored with one color. So it's not possible for R' to have an equal amount of black and white area.

WLOG, suppose $a - \lfloor a \rfloor > \frac{1}{2}$. Then R' is as follows: (the first one is for $b - \lfloor b \rfloor \leq \frac{1}{2}$, and the second one is for $b - \lfloor b \rfloor > \frac{1}{2}$)



In both cases, one may check that the black area and the white area are not the same. For example if we denote $\alpha = a - \lfloor a \rfloor$, $\beta = b - \lfloor b \rfloor$, then in the first picture,

$$\text{black area} - \text{white area} = \beta \cdot \frac{1}{2} - \beta \cdot \left(\alpha - \frac{1}{2}\right) = \beta\,(1 - \alpha) \neq 0. \tag{15.129}$$

In the second case,

$$\begin{aligned} \text{black area} - \text{white area} &= \frac{1}{4} + \left(\alpha - \frac{1}{2}\right)\left(\beta - \frac{1}{2}\right) - \left(\alpha - \frac{1}{2}\right)\frac{1}{2} - \left(\beta - \frac{1}{2}\right)\frac{1}{2} && (15.130)\\ &= (1 - \alpha)\,(1 - \beta) \neq 0. && (15.131) \end{aligned}$$

Therefore, we have proved that a rectangle R has equal amount of black and white area if and only if it has at least one side-length integer.

Now, if all subrectangles R_j have at least one integer side, then each of them have equal black and white area. Since R is a disjoint union of R_j's, R also has at equal amount of black and white area. Then, by the lemma above, R has at least one integer side. ■

I am not sure how to generalize this coloring argument to higher dimensions $d > 2$, so here is an analytic proof instead.

Alternate Solution. By a repeated application of Fubini's theorem, we get that for $f \in \mathbb{R}^d$ integrable on a rectangle $R = \prod_{j=1}^{d} [a_j, b_j] \subseteq \mathbb{R}^d$,

$$\int_R f \,\mathrm{d}\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_d}^{b_d} f \;\mathrm{d}x_d \cdots \mathrm{d}x_2 \,\mathrm{d}x_1. \tag{15.132}$$

If f is "separable", i.e.

$$f(x_1, \ldots, x_d) = f_1(x_1) \cdots f_d(x_d) = \prod_{j=1}^{d} f_j(x_j), \tag{15.133}$$

then we have

$$\int_R f \, \mathrm{d}\mathbf{x} = \prod_{j=1}^{d} \int_{a_j}^{b_j} f_j \, \mathrm{d}x_j. \tag{15.134}$$

Now, let's look at the function

$$f(x_1, \ldots, x_d) = e^{2\pi i(x_1 + \cdots + x_d)} = \prod_{j=1}^{d} e^{2\pi i x_j}. \tag{15.135}$$

Then $|f| = 1$, so it's integrable over any bounded set. Now, consider the integral

$$\int_a^b e^{2\pi i x} \, \mathrm{d}x = \frac{e^{2\pi i b} - e^{2\pi i a}}{2\pi i}, \tag{15.136}$$

which is 0 if and only if $b - a$ is an integer. Therefore, for a rectangle R,

$$\int_R f \, \mathrm{d}\mathbf{x} = \prod_{j=1}^{d} \int_{a_j}^{b_j} e^{2\pi i x_j} \, \mathrm{d}x_j = 0 \iff b_j - a_j \in \mathbb{Z} \text{ for some } j. \tag{15.137}$$

Now, each R_j has at least one side with integer length, so

$$\int_{R_j} f \, \mathrm{d}\mathbf{x} = 0. \tag{15.138}$$

Since $R = \bigcup_{j=1}^{N} R_j$, and it's an almost-disjoint union, we have

$$\chi_R = \sum_{j=1}^{N} \chi_{R_j}, \tag{15.139}$$

almost everywhere. Hence,

$$\int_R f \, \mathrm{d}\mathbf{x} = \sum_{j=1}^{N} \int_{R_j} f \, \mathrm{d}\mathbf{x} = 0. \tag{15.140}$$

So R has a side with integer length. ■

Alternate Solution. The same one as before, just replace $e^{2\pi i x}$ with $\sin 2\pi x$. ■

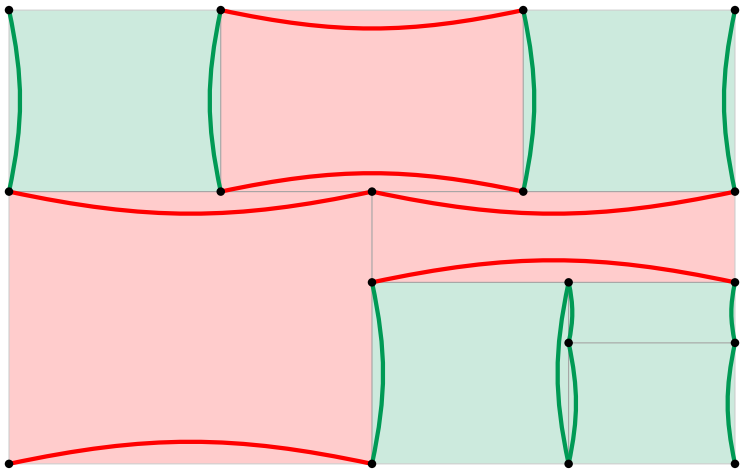
Remark 15.7. The first solution is essentially considering the function

$$f(x, y) = (-1)^{\lfloor 2x \rfloor} (-1)^{\lfloor 2y \rfloor}. \tag{15.141}$$

Although the proof is combinatorial in nature, using this analytic flavour, we can generalize it to higher dimensions. I have another combinatorial proof (below), which I don't know how to

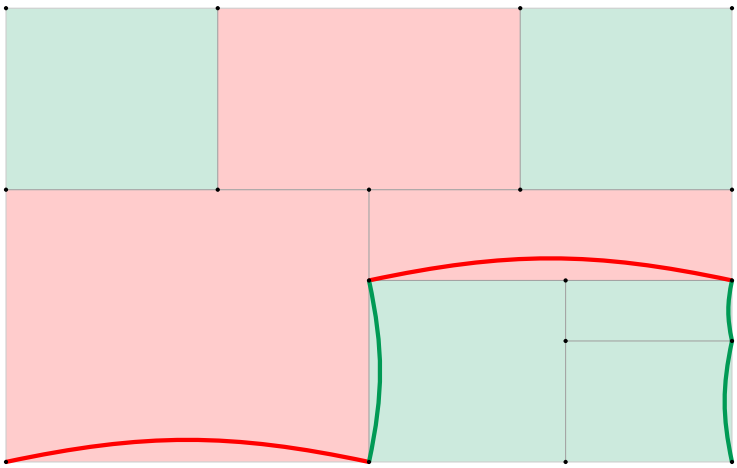
generalize to higher dimensions.

Alternate Solution to Exercise 15.3. Again, WLOG, suppose the rectangle is $[0, a] \times [0, b]$ for $a, b > 0$. We construct a graph, whose vertices are the vertices of the R_i's. We call a rectangle red if its width is an integer, and green if its height is an integer (we break ties arbitrarily). The edges of our graph are top and bottom edges of red rectangles, and left and right edges of green rectangles. We allow multi-edges.



Other than the corners of R, every vertex is in 2 or 4 rectangles. Each rectangle is contributing $+1$ to each of its vertices' degrees. Therefore, those non-corner vertices have degree 2 or 4. The corners have degree 1.

Now, let's start a path on the corner $(0, 0)$. The path will end at a vertex of degree 1, i.e. another corner.



If the path ends at $(a, 0)$, then we have travelled in the horizontal direction across red edges, which have integer lengths. Therefore, a is an integer. Similarly, if the path ends at $(0, b)$, then $b \in \mathbb{Z}$. On the other hand, if the path ends at (a, b), then both $a, b \in \mathbb{Z}$. ■

§16. 15 October 2026

The goal of today's discussion is to study differentiability of functions defined through integrals.

§16.i. The fundamental theorem of calculus

The Fundamental Theorem of Calculus essentially says that derivative and integral are inverse operations, in the same way that addition and subtraction, multiplication and division, or exponentiation and taking logarithms are opposites.

Theorem 16.1 (Fundamental theorem of calculus)

Let f be continuous on the interval $[a, b]$ and define

$$F(x) = \int_a^x f(y)\,\mathrm{d}y. \tag{16.1}$$

Then:

(i) $F'(x) = f(x)$ for every $x \in [a, b]$, i.e. the derivative of the integral goes back to the function;

(ii) if G is a C^1 function, then for $c < d$

$$\int_c^d G'(x)\,\mathrm{d}x = G(d) - G(c), \tag{16.2}$$

i.e. the integral of the derivative goes back to the function (up to a constant).

If f is continuous, then F is a C^1 function. The goal is to understand these two statements under weaker regularity assumptions on the function. In this class we focus on the first statement (differentiation of the integral); in the next class we focus on the second.

For the first statement, we compute

$$\frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f(y)\,\mathrm{d}y = \lim_{h\to 0}\frac{F(x+h)-F(x)}{h} = \lim_{h\to 0}\frac{1}{h}\int_x^{x+h} f(y)\,\mathrm{d}y. \tag{16.3}$$

The right-hand side is an *average* of f over the interval from x to $x+h$: we integrate and divide by the size of the interval. When $h \to 0$, the average over a smaller and smaller interval around the fixed point x should converge to $f(x)$. This is reasonable since f is continuous.

The theorem we are going to prove is the same statement for f merely integrable (not continuous). Of course we cannot expect convergence at *every* point, so the best we can hope for is convergence *almost everywhere*.

Definition 16.1 (Locally integrable functions). A measurable function $f : \mathbb{R}^d \to \mathbb{R}$ is called **locally integrable** if it is in $L^1(B)$ for every ball B (or, equivalently, if it's in $L^1(K)$ for

every compact subset $K \subseteq \mathbb{R}^d$). The space of all locally integrable functions is denoted by $L^1_{\text{loc}}(\mathbb{R}^d)$.

Remark 16.1. Of course $L^1\left(\mathbb{R}^d\right) \subset L^1_{\text{loc}}\left(\mathbb{R}^d\right)$. This containment is strict, because a locally integrable function may fail to be intregrable. For example, $f : \mathbb{R} \to \mathbb{R}, x \mapsto 1$ is not integrable, since

$$\int_{\mathbb{R}} |f(x)| \, \mathrm{d}x = \int_{\mathbb{R}} 1 \, \mathrm{d}x = \infty, \tag{16.4}$$

whereas

$$\int_{(a,b)} |f(x)| \, \mathrm{d}x = \int_{(a,b)} 1 \, \mathrm{d}x = b - a < \infty. \tag{16.5}$$

Theorem 16.2 (Lebesgue differentiation theorem)

Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. Then

$$\lim_{r \to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} f(y) \, \mathrm{d}y = f(x) \tag{16.6}$$

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

Here we fix a point x, take the average of f over the ball $B(x,r)$ centered at x with radius r, and let the radius shrink. Some people write the average using the notation

$$⨍_{B(x,r)} f(y) \, \mathrm{d}y := \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} f(y) \, \mathrm{d}y. \tag{16.7}$$

Lebesgue differentiation theorem is the first analogue of the fundamental theorem of calculus for locally integrable functions.

Remark 16.2. To prove a pointwise convergence result in analysis, one usually needs two ingredients:

1. the result is true for a large (dense) subclass of functions;
2. a boundedness result for an associated *maximal operator*.

Here, Lebesgue differentiation theorem is certainly true when f is continuous, which is a dense class of functions. The second ingredient is the boundedness of the maximal operator defined below.

§16.ii. Hardy–Littlewood maximal function

Definition 16.2 (Hardy–Littlewood maximal function)**.** Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. The **(centered)**

Hardy–Littlewood maximal function of f is

$$Mf(x) = \sup_{r>0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y)| \, \mathrm{d}y. \tag{16.8}$$

We put the absolute value inside because we want to control the averages, and instead of one particular average we take the *worst possible* average around x. The supremum need not be attained and may be $+\infty$. The maximal operator is the object that controls the worst possible case.

We can also define the uncentered Hardy–Littlewood maximal function:

$$\widetilde{M}f(x) = \sup_{\text{balls } B \ni x} \frac{1}{m\left(B\right)} \int_{B} |f(y)| \, \mathrm{d}y. \tag{16.9}$$

The difference is that previously the balls were centered at x, and now we don't require the balls to be centered at x. Now, since we're taking the supremum over a larger class now, one can conclude that

$$Mf(x) \leq \widetilde{M}f(x). \tag{16.10}$$

In fact, one can prove the following:

Proposition 16.3

For $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$,

$$Mf(x) \leq \widetilde{M}f(x) \leq 2^d Mf(x). \tag{16.11}$$

Proof. We have already shown first inequality. Now we'll show the second inequality. Let $B\left(c,r\right)$ be a ball containing x. For $y \in B\left(c,r\right)$,

$$|y-x| = |y-c| + |c-x| < r + r = 2r, \tag{16.12}$$

so $B\left(c,r\right) \subseteq B\left(x,2r\right)$. Therefore,

$$\begin{aligned}
\frac{1}{m\left(B\left(c,r\right)\right)} \int_{B(c,r)} |f(y)| \, \mathrm{d}y &\leq \frac{1}{m\left(B\left(c,r\right)\right)} \int_{B(x,2r)} |f(y)| \, \mathrm{d}y && (16.13)\\
&= \frac{m\left(B\left(x,2r\right)\right)}{m\left(B\left(c,r\right)\right)} \frac{1}{m\left(B\left(x,2r\right)\right)} \int_{B(x,2r)} |f(y)| \, \mathrm{d}y && (16.14)\\
&= 2^d \, \frac{1}{m\left(B\left(x,2r\right)\right)} \int_{B(x,2r)} |f(y)| \, \mathrm{d}y && (16.15)\\
&\leq 2^d \, Mf(x). && (16.16)
\end{aligned}$$

Taking the supremum over all balls $B\left(c,r\right) \ni x$, we get

$$\widetilde{M}f(x) \leq 2^d Mf(x). \tag{16.17}$$

■

Lemma 16.4

Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. For a fixed $r > 0$, the function $\Phi_r(x) : \mathbb{R}^d \to \mathbb{R}$

$$x \mapsto \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y)| \, \mathrm{d}y \tag{16.18}$$

is continuous.

Proof. Let's take a sequence $x_n \to x$. We can assume WLOG that $x_n \in B\left(x, 1\right)$. Then for any $y \in B\left(x_n, r\right)$,

$$|y - x| \leq |y - x_n| + |x_n - x| < r + 1. \tag{16.19}$$

In other words, $B\left(x_n, r\right)$ is contained in $B\left(x, r+1\right)$.

$$\Phi_r\left(x_n\right) = \frac{1}{m\left(B(x_n,r)\right)} \int_{B(x_n,r)} |f(y)| \, \mathrm{d}y = \frac{1}{m\left(B(0,r)\right)} \int_{\mathbb{R}^d} |f| \, \chi_{B(x_n,r)} \, \mathrm{d}m. \tag{16.20}$$

The integrand $|f| \, \chi_{B(x_n,r)}$ is dominated by

$$|f| \, \chi_{B(x_n,r)} \leq |f| \, \chi_{B(x,r+1)} \in L^1(\mathbb{R}^d), \tag{16.21}$$

since $f \in L^1\left(B(x, r+1)\right)$ (as it's locally integrable). Therefore, by Dominated Convergence Theorem,

$$\lim_{n\to\infty} \Phi_r\left(x_n\right) = \lim_{n\to\infty} \frac{1}{m\left(B(0,r)\right)} \int_{\mathbb{R}^d} |f| \, \chi_{B(x_n,r)} \, \mathrm{d}m \tag{16.22}$$

$$= \frac{1}{m\left(B(0,r)\right)} \int_{\mathbb{R}^d} |f| \lim_{n\to\infty} \chi_{B(x_n,r)} \, \mathrm{d}m. \tag{16.23}$$

Now, $\chi_{B(x_n,r)} \to \chi_{B(x,r)}$ pointwise a.e. (it converges everywhere except for $\partial B(x,r)$, which has measure 0). Therefore,

$$\lim_{n\to\infty} \Phi_r\left(x_n\right) = \frac{1}{m\left(B(0,r)\right)} \int_{\mathbb{R}^d} |f| \, \chi_{B(x,r)} \, \mathrm{d}m = \Phi_r(x), \tag{16.24}$$

and hence Φ_r is continuous. ∎

Proposition 16.5

Let $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. Then Mf is a measurable function. Also, the superlevel sets

$$E_\alpha = \left\{ x \in \mathbb{R}^d : Mf(x) > \alpha \right\} \tag{16.25}$$

are open for each $\alpha \in \mathbb{R}$.

Proof. Each function $x \mapsto \frac{1}{m(B(x,r))} \int_{B(x,r)} |f(y)| \, \mathrm{d}y$ is measurable, and hence their supremum is also measurable!

Let $x_0 \in E_\alpha$. Then there exists some $r > 0$ such that

$$\Phi_r(x_0) > \alpha. \tag{16.26}$$

Since Φ_r is continuous (Lemma 16.4), the set

$$U = \left\{ x \in \mathbb{R}^d : \Phi_r(x) > \alpha \right\} = \Phi_r^{-1}(a, \infty) \tag{16.27}$$

is open, and it is contained in E_α. For each $x_0 \in E_\alpha$, we can find an open set containing x_0 that is contained in E_α. Therefore, E_α is open. ■

If we instead work with uncentered Hardy-Littlewood maximal operator $\widetilde{M}$ then one can straightforwardly obtain that

$$\widetilde{E}_\alpha = \left\{ x \in \mathbb{R}^d : \widetilde{M} f(x) > \alpha \right\} \tag{16.28}$$

is open. Because, For $x \in \widetilde{E}_\alpha$, there exists some ball B such that

$$\frac{1}{m(B)} \int_B |f(y)| \, \mathrm{d}y > \alpha. \tag{16.29}$$

Then for all $x' \in B$,

$$\widetilde{M} f(x') = \sup_{\text{balls } B \ni x'} \frac{1}{m(B)} \int_B |f(y)| \, \mathrm{d}y > \alpha, \tag{16.30}$$

so we have $x \in B \subseteq \widetilde{E}_\alpha$, proving that $\widetilde{E}_\alpha$ is open.

Remark 16.3. For $f \in L^1$, it need not be true that $Mf \in L^1$. Assume $f \in L^1(\mathbb{R}^d)$ is integrable on the whole space, $f \not\equiv 0$. Then there is $R > 0$ such that the ball $B(0, R)$ contains $1 - \varepsilon$ of the mass of f, i.e.

$$\int_{B(0,R)} |f| \geq (1 - \varepsilon) \|f\|_{L^1}. \tag{16.31}$$

Take a point x far away, $|x| > R$. One particular average at x is the one over the ball $B(x, |x| + R)$, which contains $B(0, R)$. Hence

$$Mf(x) \geq \frac{1}{m\left(B(x, |x| + R)\right)} \int_{B(x,|x|+R)} |f(y)| \, \mathrm{d}y \geq \frac{(1 - \varepsilon) \|f\|_{L^1}}{\omega_d \left(|x| + R\right)^d} \tag{16.32}$$

where ω_d is the volume of the unit ball. For R fixed and $|x| \to \infty$, this behaves like $\frac{c}{|x|^d}$ with c depending only on d (and on f). Since $|x|^{-d}$ is not integrable at infinity, $Mf \notin L^1(\mathbb{R}^d)$ unless $f = 0$. So there is no hope that M maps L^1 to L^1.

However, $|x|^{-d}$ is the prototype of a function which is not in L^1 but is in L^1_{weak}. So there is hope that $Mf \in L^1_{\text{weak}}$, and this is exactly the boundedness result we shall prove.

Theorem 16.6 (Weak-type $(1,1)$ estimate)

The operator M is bounded from $L^1(\mathbb{R}^d)$ to $L^1_{\text{weak}}(\mathbb{R}^d)$, i.e. there exists a constant C, depending only on the dimension d (and not on f), such that for every $\alpha > 0$

$$m\left(\left\{x \in \mathbb{R}^d : Mf(x) > \alpha\right\}\right) \leq \frac{C}{\alpha} \|f\|_{L^1}. \tag{16.33}$$

In particular, the L^1_{weak} quasi-norm of Mf is bounded by $C\|f\|_{L^1}$.

§16.iii. Weak-type $(1,1)$ estimate implies Lebesgue differentiation theorem

Suppose for the moment that we know Theorem 16.6. We prove Theorem 16.2.

Proof of Lebesgue differentiation theorem, assuming Weak-type $(1,1)$ estimate. The statement is local. To show (16.6) for almost every $x \in B(0,R)$, it suffices to look at f on a slightly larger ball, so we can replace f by

$$f \cdot \chi_{B(0,R+1)} \tag{16.34}$$

(for small radii $r < 1$ the averages at points of $B(0,R)$ only see f inside $B(0,R+1)$). Repeating this for all balls $B(0,N)$, we obtain the result almost everywhere in $\mathbb{R}^d$. Hence we may assume $f \in L^1(\mathbb{R}^d)$.

Since $f(x)$ is just a number, it can be written as the average of the constant $f(x)$ over the ball, so (16.6) is equivalent to

$$\lim_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} \left(f(y) - f(x)\right) \mathrm{d}y = 0 \quad \text{for a.e. } x. \tag{16.35}$$

We will actually show the stronger statement (with the absolute value inside)

$$\limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y = 0 \quad \text{for a.e. } x, \tag{16.36}$$

which clearly implies the previous one.

For $\alpha > 0$ define

$$E_\alpha = \left\{ x \in \mathbb{R}^d : \limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y > 2\alpha \right\}. \tag{16.37}$$

The limsup is always nonnegative, so (16.36) holds almost everywhere as soon as $m\left(E_\alpha\right) = 0$ for every $\alpha > 0$.

Fix $\varepsilon > 0$. Since continuous compactly supported functions are dense in $L^1(\mathbb{R}^d)$, choose g continuous with

$$\|f - g\|_{L^1} < \varepsilon. \tag{16.38}$$

By the triangle inequality,

$$|f(y) - f(x)| \leq |f(y) - g(y)| + |g(y) - g(x)| + |g(x) - f(x)|\,. \tag{16.39}$$

Averaging over $B(x,r)$ (this is allowed for every ball) and taking $\limsup_{r\to 0}$, we get

$$\limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y) - f(x)|\, \mathrm{d}y \leq I_1(x) + I_2(x) + I_3(x), \tag{16.40}$$

where

$$I_1(x) = \limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y) - g(y)|\, \mathrm{d}y, \tag{16.41}$$

$$I_2(x) = \limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |g(y) - g(x)|\, \mathrm{d}y, \tag{16.42}$$

$$I_3(x) = \limsup_{r\to 0} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |g(x) - f(x)|\, \mathrm{d}y. \tag{16.43}$$

We estimate each term:

- $I_2(x) = 0$ for every x, because g is continuous.
- $I_1(x) \leq M(f-g)(x)$, since the limsup as $r \to 0$ is at most the supremum over all $r > 0$.
- $I_3(x) = |g(x) - f(x)|$, since we integrate a constant in y.

Define

$$A_\alpha = \{x : M(f-g)(x) > \alpha\}\,, \qquad B_\alpha = \{x : |g(x) - f(x)| > \alpha\}\,. \tag{16.44}$$

If $x \in E_\alpha$, then $I_1(x) + I_2(x) + I_3(x) > 2\alpha$, and since $I_2 = 0$, one of the two remaining terms must be larger than α. Therefore

$$E_\alpha \subseteq A_\alpha \cup B_\alpha. \tag{16.45}$$

By Weak-type $(1,1)$ estimate applied to $f - g$,

$$m\left(A_\alpha\right) \leq \frac{C}{\alpha} \|f - g\|_{L^1} \leq \frac{C\varepsilon}{\alpha}. \tag{16.46}$$

By Chebyshev's inequality,

$$m\left(B_\alpha\right) \leq \frac{1}{\alpha} \|f - g\|_{L^1} \leq \frac{\varepsilon}{\alpha}. \tag{16.47}$$

Hence

$$m\left(E_\alpha\right) \leq m\left(A_\alpha\right) + m\left(B_\alpha\right) \leq \frac{(C+1)\varepsilon}{\alpha}. \tag{16.48}$$

The left-hand side does not depend on g or ε, and $\varepsilon > 0$ is arbitrary, so $m\left(E_\alpha\right) = 0$. This proves Lebesgue differentiation theorem, assuming Weak-type $(1,1)$ estimate. ∎

Remark 16.4. Note that we used both ingredients: the result for a dense class of functions (continuous functions), and the boundedness of the associated maximal operator, combined with Chebyshev's inequality.

§16.iv. The Vitali covering lemma

It remains to prove Weak-type $(1,1)$ estimate. The proof is a beautiful interplay between geometry and analysis. It is based on a basic geometric fact.

Lemma 16.7 (Vitali covering lemma)

Let $B_1, B_2, \dots, B_n$ be a finite collection of balls (open or closed, as you wish) in $\mathbb{R}^d$. Then there exists a disjoint subcollection $B_{i_1}, B_{i_2}, \dots, B_{i_k}$ such that

$$m\left(\bigcup_{j=1}^{k} B_{i_j}\right) \geq C\, m\left(\bigcup_{i=1}^{n} B_i\right), \tag{16.49}$$

where $C > 0$ is a universal constant depending only on the dimension d (and not on the collection of balls).

Proof. The idea is simple: pick balls greedily, starting with the largest.

Choose B_{i_1} to be a ball of largest radius (if there are several, choose any). Delete all balls that intersect B_{i_1} (including B_{i_1} itself); these can no longer be chosen. The key observation is that, since B_{i_1} has the largest radius, every ball deleted at this step is contained in the ball $B_{i_1}^*$ with the same center as B_{i_1} and 3 times its radius.

Among the remaining balls (which are disjoint from B_{i_1}), choose one with the largest radius, B_{i_2}, delete all balls that touch it, and repeat. The process stops after finitely many steps because the collection is finite. We obtain $B_{i_1}, \dots, B_{i_k}$, which are pairwise disjoint by construction.

Every original ball is contained in some $B_{i_j}^*$. Therefore

$$m\left(\bigcup_{i=1}^{n} B_i\right) \leq m\left(\bigcup_{j=1}^{k} B_{i_j}^*\right) \leq \sum_{j=1}^{k} m\left(B_{i_j}^*\right) = 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) = 3^d m\left(\bigcup_{j=1}^{k} B_{i_j}\right), \tag{16.50}$$

where we used that the B_{i_j} are disjoint. So the lemma holds with $C = 3^{-d}$. ∎

The constant $C = 3^{-d}$ is tiny, but this does not matter: what matters is that it is universal and does not depend on the collection of balls.

§16.v. Proof of the weak-type $(1,1)$ estimate

Proof of Weak-type $(1,1)$ estimate. Fix $\alpha > 0$ and let

$$E_\alpha = \left\{x \in \mathbb{R}^d : Mf(x) > \alpha\right\}. \tag{16.51}$$

We want to show that $m(E_\alpha) \leq \frac{C}{\alpha} \|f\|_{L^1}$.

Firstly, by Proposition 16.5, E_α is open, and in particular, measurable. For each $x \in E_\alpha$, since $Mf(x) > \alpha$, there is a ball B_x (centered at x for the centered operator, or merely containing x for the uncentered one) such that

$$\frac{1}{m\left(B_x\right)} \int_{B_x} |f(y)| \, \mathrm{d}y > \alpha. \tag{16.52}$$

Let $K \subseteq E_\alpha$ be a compact set. Then $K \subseteq \bigcup_{x \in E_\alpha} B_x$ (we use open balls; if the balls are closed, take the open ball inside, which does not change the argument), so by compactness we can select finitely many balls $B_{x_1}, \ldots, B_{x_n}$ covering K.

By Vitali covering lemma, there are disjoint balls $B_{i_1}, \ldots, B_{i_k}$ among them with

$$m\left(\bigcup_{j=1}^{k} B_{i_j}\right) \geq 3^{-d} m\left(\bigcup_{l=1}^{n} B_{x_l}\right) \geq 3^{-d} m\left(K\right). \tag{16.53}$$

Using (16.52) on each of these balls, which are disjoint,

$$\|f\|_{L^1} \geq \sum_{j=1}^{k} \int_{B_{i_j}} |f(y)| \, \mathrm{d}y \geq \alpha \sum_{j=1}^{k} m\left(B_{i_j}\right) = \alpha m\left(\bigcup_{j=1}^{k} B_{i_j}\right) \geq \alpha \, 3^{-d} m\left(K\right). \tag{16.54}$$

Hence $m\left(K\right) \leq \frac{3^d}{\alpha} \|f\|_{L^1}$.

The compact set $K \subseteq E_\alpha$ was arbitrary, and the measure of the open set E_α is the supremum of the measures of its compact subsets (inner regularity, Corollary 6.5). Therefore

$$m\left(E_\alpha\right) \leq \frac{3^d}{\alpha} \|f\|_{L^1}, \tag{16.55}$$

which is (16.33) with $C = 3^d$. ∎

Remark 16.5. For $f \in L^1_{\mathrm{loc}}\left(\mathbb{R}^d\right)$, the **Lebesgue set** of f is the set of points x such that $f(x)$ is finite and

$$\lim_{r \to 0^+} \frac{1}{m\left(B(x,r)\right)} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y = 0. \tag{16.56}$$

The proof of Lebesgue differentiation theorem shows that almost every point in $\mathbb{R}^d$ is in the Lebesgue set of f.

Remark 16.6. The constant 3^d is huge, but universal. The only thing that matters for us is the universality, not the size of the constant. The conclusion is that the measure of the superlevel set of the maximal function is bounded by a constant times the L^1 norm of the function divided by α.

Remark 16.7. Note that, all the proofs we have done work verbatim if we replace the centered Hardy–Littlewood maximal operator M by the uncentered one $\widetilde{M}$. Not only that, all the results

continue to hold if we replace the balls $B(x,r)$ by a family of sets $\mathcal{S}$ that contracts around x in a regular way, i.e.

(i) the diameters of the sets in $\mathcal{S}$ go to zero;

(ii) there is a constant C such that, for each $U \in \mathcal{S}$, we have a ball B of center x so that $U \subseteq B$ and $m(B) \leq C\, m(U)$.

It is worth mentioning that the sets in the family $\mathcal{S}$ do not necessarily need to contain the point x. For such a family, we define the maximal operator as

$$M_{\mathcal{S}} f(x) := \sup_{U \in \mathcal{S}} \frac{1}{m(U)} \int_U |f(y)|\, \mathrm{d}y. \tag{16.57}$$

Then we have

$$\frac{1}{m(U)} \int_U |f(y)|\, \mathrm{d}y \leq \frac{1}{m(U)} \int_B |f(y)|\, \mathrm{d}y \leq \frac{1}{C\, m(B)} \int_B |f(y)|\, \mathrm{d}y \leq C'\, Mf(x), \tag{16.58}$$

for $C' = \frac{1}{C}$. So, by taking supremum over $U \in \mathcal{S}$, we have

$$M_{\mathcal{S}} f(x) \leq C'\, Mf(x). \tag{16.59}$$

As a result,

$$E_{\alpha,\mathcal{S}} = \left\{ x \in \mathbb{R}^d : M_{\mathcal{S}} f(x) > \alpha \right\} \subseteq E_{\frac{\alpha}{C'}} = \left\{ x \in \mathbb{R}^d : Mf(x) > \frac{\alpha}{C'} \right\}. \tag{16.60}$$

So we have a Weak-type $(1,1)$ estimate for this family $\mathcal{S}$ as well, namely,

$$m\left(\left\{ x \in \mathbb{R}^d : M_{\mathcal{S}} f(x) > \alpha \right\} \right) \leq m\left(\left\{ x \in \mathbb{R}^d : Mf(x) > \frac{\alpha}{C'} \right\} \right) \leq \frac{CC'}{\alpha} \|f\|_{L^1}, \tag{16.61}$$

where the constant $C'' := CC'$ depends only on d. Then the proof of Lebesgue differentiation theorem works word-by-word if we replace balls centered at x by this family $\mathcal{S}$, and M by $M_{\mathcal{S}}$.

Remark 16.8 (The same framework elsewhere)**.** We saw the following framework for proving a pointwise convergence result: it must hold for a dense class of functions, and one needs a boundedness estimate for an associated maximal operator. The same philosophy appears many times in analysis:

- Convergence of the solution of a PDE (for instance in the upper half-space) to its initial datum as time goes to zero: one proves it for a dense class of functions and controls the associated maximal operator.
- Pointwise convergence of Fourier series: true for Schwartz functions, and for L^2 one needs the Carleson maximal operator, shown by Carleson in the 1960s to be bounded from L^2 to weak L^2 (in the sense of Carleson's theorem, later extended by others).
- Ergodic averages in dynamical systems: one controls the associated maximal operator.

§16.vi. Exercises

Exercise 16.1. Let M denote the centered Hardy-Littlewood maximal operator. Suppose $d \geq 3$, and consider the function

$$f_\alpha(x) = |x|^{-\alpha} \tag{16.62}$$

for $0 < \alpha < d$.

(a) Show that

$$Mf_\alpha(x) = C_{\alpha,d} f_\alpha(x), \tag{16.63}$$

for all $x \in \mathbb{R}^d$ where $C_{\alpha,d}$ is a constant (that may depend on α and d).

(b) For each α, with $0 < \alpha < d$, decide if $C_{\alpha,d} = 1$ or $C_{\alpha,d} > 1$. Justify.

Solution. Throughout, we will write $f := f_\alpha$, $C := C_{\alpha,d}$.

(a) First, it's obviously true for $x = 0$, because $f_\alpha(0) = +\infty$, and

$$Mf(0) = \sup_{R>0} \frac{1}{m(B(0,R))} \int_{B(0,R)} |y|^{-\alpha} \, \mathrm{d}y = \sup_{R>0} \frac{1}{\omega_d R^d} d\omega_d \int_0^R r^{d-1} r^{-\alpha} \, \mathrm{d}r, \tag{16.64}$$

in polar coordinates. The integrand is $r^{d-1-\alpha}$, which integrates to $\frac{r^{d-\alpha}}{d-\alpha}$, hence,

$$Mf(0) = \sup_{R>0} \frac{d}{R^d} \frac{R^{d-\alpha}}{d-\alpha} = \sup_{R>0} \frac{d}{d-\alpha} R^{-\alpha} = +\infty \tag{16.65}$$

since $\alpha > 0$, so taking $R \to 0^+$ leads to infinity.

So now we consider $x \neq 0$. We have

$$Mf(x) = \sup_{R>0} \frac{1}{m(B(x,R))} \int_{B(x,R)} |y|^{-\alpha} \, \mathrm{d}y = \sup_{R>0} \frac{1}{\omega_d R^d} \int_{B(x,R)} |y|^{-\alpha} \, \mathrm{d}y. \tag{16.66}$$

We write $R = |x| \lambda$. Then a supremum over $R > 0$ will be a supremum over $\lambda > 0$. Also, we'll do a change of variables $y = |x| z$, so that $\mathrm{d}y = |x|^d \, \mathrm{d}z$. Then the domain of integration will be

$$|x| z \in B(x,R) \iff z \in B\left(\frac{x}{|x|}, \frac{R}{|x|}\right) = B(\widehat{x}, \lambda), \tag{16.67}$$

where $\widehat{x} := \frac{x}{|x|} \in S^{d-1}$ is the unit vector in the direction of $x \in \mathbb{R}^d$. Then

$$Mf(x) = \sup_{\lambda>0} \frac{1}{\omega_d |x|^d \lambda^d} \int_{B(\widehat{x},\lambda)} |x|^{-\alpha} |z|^{-\alpha} \, |x|^d \, \mathrm{d}z \tag{16.68}$$

$$= |x|^{-\alpha} \sup_{\lambda>0} \frac{1}{\omega_d \lambda^d} \int_{B(\widehat{x},\lambda)} |z|^{-\alpha} \, \mathrm{d}z. \tag{16.69}$$

The integrand is radially symmetric. So the integral stays the same if we choose any other center in S^{d-1} other than $\widehat{x}$, so let's take $e_1 = (1, 0, \ldots, 0)$ to be the center, and denote

$$C := \sup_{\lambda>0} \frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z. \tag{16.70}$$

We claim that this is our desired constant! First, we need to show it's finite. For $\lambda \leq \frac{1}{2}$, the domain $B(e_1, \lambda)$ doesn't contain the origin. In fact, for $z \in B(e_1, \lambda)$,

$$|z| \geq |e_1| - |z - e_1| > 1 - \frac{1}{2} = \frac{1}{2} \implies |z|^{-\alpha} \leq 2^\alpha. \tag{16.71}$$

Therefore,

$$\frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z. \leq \frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} 2^\alpha \, \mathrm{d}z = 2^\alpha. \tag{16.72}$$

For $\lambda > \frac{1}{2}$, the integrating region $B(e_1, \lambda)$ is contained in $B(0, \lambda + 1)$. So

$$\begin{aligned} \frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z &\leq \frac{1}{\omega_d \lambda^d} \int_{B(0,\lambda+1)} |z|^{-\alpha} \, \mathrm{d}z & (16.73) \\ &= \frac{1}{\omega_d \lambda^d} \left[\frac{d}{d-\alpha} \omega_d (1+\lambda)^{d-\alpha} \right] & (16.74) \\ &= \frac{d}{d-\alpha} \frac{(1+\lambda)^{d-\alpha}}{\lambda^d}, & (16.75) \end{aligned}$$

where the integral $\int_{B(0,\lambda+1)} |z|^{-\alpha} \, \mathrm{d}z$ is obtained from the $x = 0$ case, (16.64), (16.65). Now, the function

$$\lambda \mapsto \frac{(1+\lambda)^{d-\alpha}}{\lambda^d} \tag{16.76}$$

is continuous on $\left[\frac{1}{2}, \infty\right)$, and

$$\lim_{\lambda\to\infty} \frac{(1+\lambda)^{d-\alpha}}{\lambda^d} = \lim_{\lambda\to\infty} \frac{(1+\lambda)^{d-\alpha}}{\lambda^{d-\alpha}} \cdot \lim_{\lambda\to\infty} \lambda^{-\alpha} = 1 \cdot 0 = 0 \tag{16.77}$$

Therefore, the function is bounded on $\left[\frac{1}{2}, \infty\right)$, say by M_0. Hence,

$$C \leq \max\left\{ 2^\alpha, \frac{d}{d-\alpha} M_0 \right\} < \infty. \tag{16.78}$$

(b) By Lebesgue differentiation theorem,

$$\lim_{\lambda\to 0} \frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z = f(e_1) = 1. \tag{16.79}$$

So we have

$$C = \sup_{\lambda>0} \frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z \geq 1. \tag{16.80}$$

Suppose there exists some $\beta \in (0, d)$ such that

$$\frac{1}{\omega_d \lambda^d} \int_{B(e_1,\lambda)} |z|^{-\beta} \, \mathrm{d}z \leq 1 \text{ for every } \lambda > 0 \text{ with equality for at least one } \lambda. \tag{16.81}$$

Then we claim that $C = 1$ for $0 < \alpha \leq \beta$ and $C > 1$ for $\alpha > \beta$. In the first case, $0 < \alpha \leq \beta$, consider $p = \frac{\alpha}{\beta} \in (0, 1]$. The map

$$\begin{aligned} \varphi : \mathbb{R} &\to \mathbb{R} \\ t &\mapsto t^p = t^{\frac{\alpha}{\beta}} \end{aligned} \tag{16.82}$$

is concave, since $\varphi''(t) = p(p-1)t^{p-2} \leq 0$. Therefore, by Jensen's inequality for concave functions (along with the construction from Corollary 11.6),

$$\begin{aligned} \frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z &= \frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} \varphi\left(|z|^{-\beta}\right) \, \mathrm{d}z && (16.83) \\ &\leq \varphi\left(\frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} |z|^{-\beta} \, \mathrm{d}z\right) && (16.84) \\ &\leq \varphi(1) = 1. && (16.85) \end{aligned}$$

By taking the supremum over λ, we have $C \leq 1$. Hence, for $0 < \alpha \leq \beta$, $C = 1$.

Now, for $\alpha > \beta$, the map $\varphi(t) = t^p = t^{\frac{\alpha}{\beta}}$ is strictly convex, since $\varphi''(t) = p(p-1)t^{p-2} > 0$. Combining Corollary 11.6 and Jensen's inequality for stricly convex functions, we have

$$\begin{aligned} \frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} |z|^{-\alpha} \, \mathrm{d}z &= \frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} \varphi\left(|z|^{-\beta}\right) \, \mathrm{d}z && (16.86) \\ &> \varphi\left(\frac{1}{m\left(B\left(e_1, \lambda\right)\right)} \int_{B(e_1,\lambda)} |z|^{-\beta} \, \mathrm{d}z\right) && (16.87) \\ &= \varphi(1) = 1. && (16.88) \end{aligned}$$

By taking the supremum over λ, we have $C > 1$. Hence, for $\beta < \alpha < d$, $C > 1$.

Now, it remains to show whether such a β exists so as to satisfy (16.81).

Claim 16.1: Let $U \subseteq \mathbb{R}^d$ be open, and $u \in C^2(U)$. If $u : U \to \mathbb{R}$ is a harmonic function, i.e. $\nabla^2 u = 0$, then for any $x \in U$ and $\overline{B\left(x, R\right)} \subseteq U$,

$$\frac{1}{m\left(B(x, R)\right)} \int_{B(x,R)} u(y) \, \mathrm{d}y = u(x). \tag{16.89}$$

Proof. We shall use polar coordinates centered at x.

$$\int_{B(x,R)} F\left(y\right) \, \mathrm{d}y = \int_0^R \int_{S^{d-1}} r^{d-1} F\left(x + r\,\widehat{\boldsymbol{\theta}}\right) \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}r. \tag{16.90}$$

Given $r \in [0, R]$, let $\varphi(r)$ be the spherical average of u:

$$\varphi(r) = \frac{1}{\omega_d d} \int_{S^{d-1}} u\left(x + r\widehat{\boldsymbol{\theta}}\right) \, \mathrm{d}\boldsymbol{\theta}, \tag{16.91}$$

where $\omega_d d$ is the $(d-1)$-dimensional measure of S^{d-1}. If we can prove that $\varphi(r) = u(x)$ for all $r \in [0, R]$, then we'll be done, since

$$\fint_{B(x,R)} u(y)\,\mathrm{d}y = \frac{1}{\omega_d R^d} \int_0^R r^{d-1} \left(\int_{S^{d-1}} u\left(x + r\,\widehat{\boldsymbol{\theta}}\right) \, \mathrm{d}\boldsymbol{\theta} \right) \mathrm{d}r \tag{16.92}$$

$$= \frac{1}{\omega_d R^d} \int_0^R r^{d-1} \, d\omega_d \, u(x)\,\mathrm{d}r \tag{16.93}$$

$$= u(x). \tag{16.94}$$

So we'll prove that $\varphi(r) = u(x)$ for all $r \in [0, R]$. First, let's show that φ is continuous. u is continuous on the compact set $\overline{B(x, R)}$, so it's uniformly continuous. Therefore, given $\varepsilon > 0$, there exists $\delta > 0$ such that whenever $|x_1 - x_2| < \delta$, $|u(x_1) - u(x_2)| \leq \varepsilon$. Hence, for $|r - r'| < \delta$,

$$\left| u\left(x + r\widehat{\boldsymbol{\theta}}\right) - u\left(x + r'\widehat{\boldsymbol{\theta}}\right) \right| < \varepsilon, \tag{16.95}$$

since $\left| x + r\widehat{\boldsymbol{\theta}} - x - r'\widehat{\boldsymbol{\theta}} \right| = |r - r'| < \delta$. As a result,

$$\left|\varphi(r) - \varphi\left(r'\right)\right| \leq \frac{1}{\omega_d d} \int_{S^{d-1}} \left| u\left(x + r\widehat{\boldsymbol{\theta}}\right) - u\left(x + r'\widehat{\boldsymbol{\theta}}\right) \right| \mathrm{d}\boldsymbol{\theta} \tag{16.96}$$

$$\leq \frac{1}{\omega_d d} \int_{S^{d-1}} \varepsilon \,\mathrm{d}\boldsymbol{\theta} = \varepsilon. \tag{16.97}$$

Hence, φ is uniformly continuous! Since u and ∇u are continuous and bounded in the integrating region, we can differentiate under the integral sign!

$$\varphi'(r) = \frac{1}{\omega_d d} \int_{S^{d-1}} \nabla u\left(x + r\widehat{\boldsymbol{\theta}}\right) \cdot \widehat{\boldsymbol{\theta}} \, \mathrm{d}\boldsymbol{\theta} = \frac{1}{\omega_d d r^{d-1}} \int_{\partial B(x,r)} \nabla u\left(y\right) \cdot \frac{y - x}{r} \tag{16.98}$$

$$= \frac{1}{\omega_d d r^{d-1}} \int_{B(x,r)} \nabla^2 u \,\mathrm{d}y = 0, \tag{16.99}$$

by the Divergence theorem. Now, $\varphi' = 0$, so it's constant on $(0, R)$. By continuity, it's constant on $[0, R]$. By the uniform continuity of u, given $\varepsilon > 0$, there exists $\delta > 0$ such that $|r| < \delta$ implies

$$\left| u\left(x + r\widehat{\boldsymbol{\theta}}\right) - u(x) \right| < \varepsilon, \tag{16.100}$$

since $\left| x + r\widehat{\boldsymbol{\theta}} - x \right| = |r| < \delta$. Therefore,

$$|\varphi(r) - u(x)| \leq \frac{1}{\omega_d d} \int_{S^{d-1}} \left| u\left(x + r\widehat{\boldsymbol{\theta}}\right) - u(x) \right| \, \mathrm{d}\boldsymbol{\theta} \tag{16.101}$$

$$\leq \frac{1}{\omega_d d} \int_{S^{d-1}} \varepsilon \, \mathrm{d}\boldsymbol{\theta} = \varepsilon. \tag{16.102}$$

Therefore, $\varphi(0) = \lim_{r\to 0} \varphi(r) = u(x)$. And we are done! □

Now let's check for which α is f a harmonic function.

$$\frac{\partial f}{\partial x_i} = \frac{\partial}{\partial x_i}\left[\sum_{i=1}^{n} x_i^2\right]^{-\frac{\alpha}{2}} = -\frac{\alpha}{2}\left|x\right|^{-\alpha-2} 2x_i = -\alpha x_i \left|x\right|^{-\alpha-2}. \tag{16.103}$$

$$\frac{\partial^2 f}{\partial x_i^2} = -\alpha\left|x\right|^{-\alpha-2} + \alpha\left(\alpha+2\right) x_i^2 \left|x\right|^{-\alpha-4}. \tag{16.104}$$

Taking the sum over $i = 1$ to d, we have

$$\nabla^2 f = -\alpha d\left|x\right|^{-\alpha-2} + \alpha\left(\alpha+2\right)\left|x\right|^2\left|x\right|^{-\alpha-4} = \alpha\left(\alpha+2-d\right)\left|x\right|^{-\alpha-2}, \tag{16.105}$$

which is 0 if and only if $\alpha = d-2$ and $x \neq 0$. We claim that the β with the property (16.81) is $d-2$. Indeed, for $\lambda < 1$, the ball $B\left(e_1, \lambda\right)$ doesn't contain the origin, so $|z|^{d-2}$ is harmonic on $\overline{B\left(e_1, \lambda\right)}$. Therefore, by Claim 16.1,

$$\frac{1}{\omega_d \lambda^d}\int_{B(e_1,\lambda)} |z|^{-(d-2)}\, \mathrm{d}z = |e_1|^{-(d-2)} = 1 \text{ for all } 0 < \lambda < 1. \tag{16.106}$$

Now consider the case $\lambda \geq 1$. Let's denote

$$A_\lambda\left(e_1\right) := \frac{1}{\omega_d \lambda^d}\int_{B(e_1,\lambda)} |z|^{-(d-2)}\, \mathrm{d}z \tag{16.107}$$

$$= \frac{d}{\lambda^d}\int_0^\lambda r^{d-1}\frac{1}{\omega_d d}\int_{S^{d-1}} |e_1 + r\boldsymbol{\theta}|^{-(d-2)}\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}r. \tag{16.108}$$

We know that the inner integral is 1 for $0 < r < 1$ (from the proof of Claim 16.1). For $\lambda > 1$, we can write the integrand as:

$$|ae_1 + r\boldsymbol{\theta}|^{-(d-2)} = \left[a^2 + r^2 + 2ar\cos\theta\right]^{-\frac{d-2}{2}}, \tag{16.109}$$

which is symmetric in a and r. The integrand was $(a, r) = (1, r)$. Therefore, by symmetry, the integrand is equal to the case when $(a, r) = (r, 1)$. Hence, the inner integral is

$$\frac{1}{\omega_d d}\int_{S^{d-1}} |re_1 + 1\boldsymbol{\theta}|^{-(d-2)}\, \mathrm{d}\boldsymbol{\theta}, \tag{16.110}$$

which is the average over the sphere centered at re_1 with radius 1. For $r > 1$, the ball $\overline{B\left(re_1, 1\right)}$ doesn't contain 0, so $|z|^{-(d-2)}$ is harmonic here. Hence, the inner integral in (16.110) is $r^{-(d-2)} \leq 1$. As a result, from (16.108),

$$A_\lambda(e_1) \leq \frac{d}{\lambda^d}\int_0^\lambda r^{d-1}\, \mathrm{d}r = 1. \tag{16.111}$$

Therefore, $A_\lambda(e_1) \leq 1$ for all $\lambda > 0$, with equality for $\lambda < 1$. Hence, $\beta = d-2$ satisfies

(16.81). As a result,

$$C_{\alpha,d} = \begin{cases} = 1 & \text{if } 0 < \alpha \leq d-2 \\ > 1 & \text{if } d-2 < \alpha < d. \end{cases} \tag{16.112}$$

■

§17. 29 October 2026

§17.i. Functions of bounded variation

Definition 17.1 (Bounded variation). A function $F : [a, b] \to \mathbb{R}$ has **bounded variation** if

$$\operatorname{Var}_F [a, b] := \sup_{\mathcal{P}} \sum_{j=1}^{N} |F(t_j) - F(t_{j-1})| < \infty, \tag{17.1}$$

where the supremum is taken over all partitions $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_N = b\}$ of the interval $[a, b]$. The number $\operatorname{Var}_F [a, b]$ is called the **variation** of F on $[a, b]$.

In words: we take any partition of the interval, measure the "jumps" of F between consecutive points of the partition (from t_{j-1} to t_j the function increased or decreased by some amount), and sum the absolute values of these jumps. If this quantity stays bounded over all partitions, the function has bounded variation.

Remark 17.1 (First observations).

1. If a partition $\mathcal{P}'$ refines $\mathcal{P}$, then the sum in (17.1) over $\mathcal{P}'$ is greater than or equal to the sum over $\mathcal{P}$. Indeed, if we add a point t between t_{j-1} and t_j, the triangle inequality gives

$$|F(t_j) - F(t_{j-1})| \le |F(t_j) - F(t)| + |F(t) - F(t_{j-1})|. \tag{17.2}$$

   Hence, when dealing with partitions, we may always assume that the partition is as fine as we like.

2. Throughout, F is *increasing* if $F(x) \le F(y)$ for $x \le y$ (this means non-decreasing). If F is increasing and bounded, the sums in (17.1) telescope, so

$$\operatorname{Var}_F [a, b] = F(b) - F(a). \tag{17.3}$$

3. If F and G have bounded variation, then so does $F + G$, and by the triangle inequality, for every partition,

$$|(F + G)(t_j) - (F + G)(t_{j-1})| \le |F(t_j) - F(t_{j-1})| + |G(t_j) - G(t_{j-1})|, \tag{17.4}$$

   so $\operatorname{Var}_{F+G} [a, b] \le \operatorname{Var}_F [a, b] + \operatorname{Var}_G [a, b]$.

4. For $a \le x \le y \le b$ we may always include the point x in a partition of $[a, y]$, and therefore

$$\operatorname{Var}_F [a, y] = \operatorname{Var}_F [a, x] + \operatorname{Var}_F [x, y]. \tag{17.5}$$

   Also, taking the trivial partition $\{a, x\}$, we have $|F(x) - F(a)| \le \operatorname{Var}_F [a, x]$; more generally $|F(y) - F(x)| \le \operatorname{Var}_F [x, y]$.

Increasing functions are easy to handle, and the following proposition shows that they are essentially all we need.

Proposition 17.1

A function $F : [a,b] \to \mathbb{R}$ has bounded variation if and only if F can be written as a difference of two bounded and increasing functions.

Proof. If F is the difference of two bounded increasing functions, then F has bounded variation by the observations above (each increasing function has finite variation, and we use the triangle inequality).

Conversely, let F have bounded variation. Then $x \mapsto \operatorname{Var}_F [a,x]$ is increasing and bounded by $\operatorname{Var}_F [a,b]$, and we write

$$F(x) = \left(\frac{\operatorname{Var}_F [a,x] + F(x)}{2} \right) - \left(\frac{\operatorname{Var}_F [a,x] - F(x)}{2} \right). \tag{17.6}$$

Both functions are bounded. They are increasing: let $x < y$. By (17.5),

$$\operatorname{Var}_F [a,y] - \operatorname{Var}_F [a,x] = \operatorname{Var}_F [x,y] \geq |F(y) - F(x)|, \tag{17.7}$$

so that $\operatorname{Var}_F [a,y] \mp F(y) \geq \operatorname{Var}_F [a,x] \mp F(x)$. ∎

The bottom line is that, to study the differentiability of functions of bounded variation, we may restrict ourselves to increasing functions. Conversely, sums and differences of functions of bounded variation again have bounded variation.

§17.ii. Dini numbers and the main theorem

Definition 17.2 (Dini numbers)**.** Let $F : [a,b] \to \mathbb{R}$ be a bounded and increasing function. For each point $x_0 \in (a,b)$ we define the four **Dini numbers**

$$D_1 F(x_0) := \limsup_{h \to 0^+} \frac{F(x_0+h) - F(x_0)}{h}, \tag{17.8}$$

$$D_2 F(x_0) := \liminf_{h \to 0^+} \frac{F(x_0+h) - F(x_0)}{h}, \tag{17.9}$$

$$D_3 F(x_0) := \limsup_{h \to 0^-} \frac{F(x_0+h) - F(x_0)}{h}, \tag{17.10}$$

$$D_4 F(x_0) := \liminf_{h \to 0^-} \frac{F(x_0+h) - F(x_0)}{h}. \tag{17.11}$$

Here $h \to 0^+$ means that h tends to zero from the right, and $h \to 0^-$ from the left. Observe that one always has $D_1 F(x_0) \geq D_2 F(x_0)$ and $D_3 F(x_0) \geq D_4 F(x_0)$, and that if these four

numbers are equal (and finite), then F is differentiable at x_0. Since F is increasing, all of them are non-negative. We can also rewrite D_3 and D_4 as

$$D_3 F(x_0) = \limsup_{h \to 0^+} \frac{F(x_0) - F(x_0 - h)}{h}, \tag{17.12}$$

$$D_4 F(x_0) = \liminf_{h \to 0^+} \frac{F(x_0) - F(x_0 - h)}{h}. \tag{17.13}$$

Theorem 17.2

Let $F : [a, b] \to \mathbb{R}$ be a bounded and increasing function. Then F is differentiable almost everywhere in (a, b). Moreover, F' is measurable, non-negative, and

$$0 \leq \int_a^b F'(y)\, \mathrm{d}y \leq F(b) - F(a). \tag{17.14}$$

Remark 17.2. By Proposition 17.1, Theorem 17.2 implies that every function of bounded variation has a classical derivative almost everywhere (the limit of the difference quotients exists). Note that the inequality in (17.14) is not necessarily an equality: functions of bounded variation need not satisfy the Fundamental theorem of calculus. For that, the function needs a bit more regularity, which leads to the class of *absolutely continuous* functions.

Also, even though the set of discontinuities of an increasing function is countable[3], continuity at a point does not give differentiability there (there are continuous functions that are nowhere differentiable). So the theorem is not trivial, and we need a covering argument.

§17.iii. A second Vitali covering lemma

One more time, a lemma of geometric character will play an important role. It is a version of the Vitali covering lemma which is different in flavour from before. Earlier, a set was covered by balls and we could extract a disjoint subcollection covering a universal proportion of it. Here, the covering has the extra property that every point is covered by arbitrarily small balls, and the conclusion is that we can cover *essentially everything*.

Definition 17.3 (Vitali covering)**.** A collection $\mathcal{B}$ of balls $\{B\}$ (open balls, or closed balls with non-empty interior) is a **Vitali covering** of a measurable set $E \subseteq \mathbb{R}^d$ if for every $x \in E$ and every $\eta > 0$ there exists a ball $B \in \mathcal{B}$ such that $x \in B$ and $m(B) < \eta$.

So not only is every point of E covered by some ball of the collection, but it is covered by balls of arbitrarily small size.

[3] https://en.wikipedia.org/wiki/Discontinuities_of_monotone_functions

Lemma 17.3 (Vitali covering, second version)

Let $E \subseteq \mathbb{R}^d$ be a measurable set of finite measure and let $\mathcal{B}$ be a Vitali covering of E. Then, for each $\delta > 0$, we can find a finite sub-collection $B_1, B_2, \ldots, B_N$ of disjoint balls in $\mathcal{B}$ such that

$$m(E) + \delta \geq \sum_{i=1}^{N} m(B_i) \geq m(E) - \delta, \tag{17.15}$$

and

$$m\left(E \cap \bigcup_{i=1}^{N} B_i\right) \geq m(E) - 2\delta. \tag{17.16}$$

Proof. Suppose $\mathcal{B}$ consists of closed balls with non-empty interior. Given $\varepsilon > 0$, for each $B \in \mathcal{B}$ consider the open ball B^* concentric with B and with radius multiplied by $(1+\varepsilon)$. The new collection is again a Vitali covering of E, and since $\varepsilon > 0$ is arbitrary, it suffices to treat the case in which all balls are open.

Step 1: Left-hand inequality of (17.15)***, and*** (17.16)**.** By the definition of Lebesgue measure, there is an open set $O \supseteq E$ with $m(O - E) < \delta$. Restrict to the balls of $\mathcal{B}$ that are entirely contained in O. This is still a Vitali covering of E: if $x \in E \subseteq O$, there is a small ball around x inside O, and the balls of $\mathcal{B}$ containing x with small enough measure lie inside it. With this restriction, any disjoint sub-collection satisfies

$$\sum_{i=1}^{N} m(B_i) = m\left(\bigcup_{i=1}^{N} B_i\right) \leq m(O) \leq m(E) + \delta, \tag{17.17}$$

which is the left inequality in (17.15).

Step 2: Right-hand inequality of (17.15)**.** It suffices to consider $\delta < m(E)$ (otherwise the right-hand side is non-positive and there is nothing to prove). We use the first Vitali covering lemma to generate the desired sub-collection, iteratively.

Choose a compact set $E' \subseteq E$ with $m(E') \geq \delta$ (inner regularity of the Lebesgue measure, Proposition 6.4). We can cover E' with a finite sub-collection of $\mathcal{B}$. By Vitali covering lemma we find an initial disjoint sub-collection $B_1, B_2, \ldots, B_{N_1}$ in $\mathcal{B}$ such that

$$\sum_{i=1}^{N_1} m(B_i) \geq C\, m\left(\bigcup_{i=1}^{N_1} B_i\right) \geq C\, m\left(E'\right) \geq C\delta, \qquad C = 3^{-d}. \tag{17.18}$$

If $\sum_{i=1}^{N_1} m(B_i) \geq m(E) - \delta$, we are done with $N = N_1$. Otherwise $\sum_{i=1}^{N_1} m(B_i) < m(E) - \delta$, and we consider the set

$$E_2 = E - \bigcup_{i=1}^{N_1} \overline{B_i}. \tag{17.19}$$

Since the boundary of a ball has measure zero, $m(E_2) \geq m(E) - \sum_{i=1}^{N_1} m(B_i) > \delta$. Moreover, the balls of $\mathcal{B}$ that are disjoint from $\bigcup_{i=1}^{N_1} \overline{B_i}$ still form a Vitali covering of E_2. (This is why it is important to work with the *closures*: if $x \in E_2$, then x lies in the open set $\mathbb{R}^d - \bigcup_i \overline{B_i}$, so there

is a little ball around x not touching the closed union, and balls of $\mathcal{B}$ containing x with small enough measure lie inside this little ball.)

We repeat the process with E_2, and so on. Each step adds disjoint balls, disjoint from the previous ones, with total measure at least $C\delta$. After repeating this process k times we have reached a finite sub-collection of disjoint balls with total measure greater than or equal to $kC\delta$, and, in the worst case, at the k-th step with $k \geq (m(E) - \delta)/(C\delta)$ (a finite number, since $m(E) < \infty$) we hit our mark $\sum_i m(B_i) \geq m(E) - \delta$.

Step 3: Proving (17.16)**.** The right inequality in (17.15) implies (17.16): write

$$\bigcup_{i=1}^{N} B_i = \left(E \cap \bigcup_{i=1}^{N} B_i\right) \cup \left(\bigcup_{i=1}^{N} B_i - E\right) \tag{17.20}$$

a disjoint union, where the second set is contained in $O - E$. Thus

$$m\left(E \cap \bigcup_{i=1}^{N} B_i\right) = m\left(\bigcup_{i=1}^{N} B_i\right) - m\left(\bigcup_{i=1}^{N} B_i - E\right) \tag{17.21}$$

$$\geq (m(E) - \delta) - \delta = m(E) - 2\delta. \tag{17.22}$$

■

§17.iv. Proof of differentiability of BV functions

We now prove Theorem 17.2. We split the proof into three steps.

Proof of Theorem 17.2. ***Step 1: each Dini number is finite almost everywhere.*** Let us show, for instance, that the set

$$E = \{x \in (a,b) : D_1 F(x) = \infty\} \tag{17.23}$$

has measure zero; the same reasoning applies to the other Dini numbers. Define

$$E_k = \{x \in (a,b) : D_1 F(x) > k\}. \tag{17.24}$$

This is a measurable set (since $D_1 F$ is measurable), and $E = \bigcap_{k=1}^{\infty} E_k$.

If $x \in E_k$, then, since $D_1 F(x)$ is a lim sup and is larger than k, there exist arbitrarily small $h > 0$ such that $[x, x+h] \subseteq (a,b)$ and

$$\frac{F(x+h) - F(x)}{h} > k. \tag{17.25}$$

So the intervals $[x, x+h]$ with this property form a Vitali covering of E_k (the intervals are as small as we want). By Lemma 17.3 (Vitali covering, second version), given $\delta > 0$, there exists a

finite disjoint collection of intervals $[x_j, x_j + h_j]$, $j = 1, 2, \ldots, N$, such that

$$F(x_j + h_j) - F(x_j) > kh_j, \qquad \sum_{j=1}^{N} h_j \geq m(E_k) - \delta. \tag{17.26}$$

Since F is increasing and the intervals are disjoint, the sum of the increments over these intervals is at most the total increment of F, hence

$$F(b) - F(a) \geq \sum_{j=1}^{N} [F(x_j + h_j) - F(x_j)] \geq k(m(E_k) - \delta), \tag{17.27}$$

and as $\delta > 0$ is arbitrary,

$$m(E_k) \leq \frac{F(b) - F(a)}{k}. \tag{17.28}$$

Since $E \subseteq E_k$ for every k, we conclude $m(E) = 0$.

Step 2: the Dini numbers coincide almost everywhere. Let us show that the set $\{x \in (a, b) : D_1F(x) > D_4F(x)\}$ has measure zero. The same logic applies to any pair of Dini numbers (recall also that $D_1 \geq D_2$ and $D_3 \geq D_4$ always); we will conclude that all four are equal, and by Step 1 finite, a.e. in (a, b), so F is differentiable a.e.

It is sufficient to verify that each set

$$A_{r,s} := \{x \in (a, b) : D_1F(x) > r > s > D_4F(x)\} \tag{17.29}$$

has measure zero, since the original set is the (countable) union of all the sets $A_{r,s}$ with $r, s \in \mathbb{Q}$, i.e.

$$\{x \in (a, b) : D_1F(x) > D_4F(x)\} = \bigcup_{r,s \in \mathbb{Q}} A_{r,s}. \tag{17.30}$$

As F is increasing, the difference quotients are non-negative, so we may assume $r > s > 0$. The measurability of $A_{r,s}$ is clear, since D_1F, D_4F are measurable.

Fix $r > s > 0$ and let $A = A_{r,s}$. Suppose, for contradiction, that $m(A) > 0$.

(a) Intervals on the left. For each $x \in A$, since $D_4F(x) < s$, there exists a sequence of $h > 0$, arbitrarily small, such that

$$\frac{F(x - h) - F(x)}{-h} < s, \quad \text{i.e.} \quad F(x) - F(x - h) < sh. \tag{17.31}$$

The intervals $[x - h, x]$ form a Vitali covering of A. By Lemma 17.3 (Vitali covering, second version), given $\delta > 0$, there exists a finite collection of disjoint intervals $[x_j - h_j, x_j]$, $j = 1, 2, \ldots, N$, such that

$$F(x_j) - F(x_j - h_j) < sh_j, \tag{17.32}$$

$$m\left(A \cap \bigcup_{j=1}^{N} [x_j - h_j, x_j]\right) \geq m(A) - 2\delta, \tag{17.33}$$

and

$$\sum_{j=1}^{N} h_j \leq m(A) + \delta. \tag{17.34}$$

Combining (17.32) and (17.34) we arrive at

$$\sum_{j=1}^{N} [F(x_j) - F(x_j - h_j)] < s(m(A) + \delta). \tag{17.35}$$

(b) Intervals on the right. Consider now the set $B = A \cap \bigcup_{j=1}^{N} [x_j - h_j, x_j]$. Discarding the finitely many endpoints (a set of measure zero), each $y \in B$ lies in the interior of one of these intervals. For each $y \in B$, the fact that $D_1 F(y) > r$ implies the existence of a sequence of $k > 0$, arbitrarily small, such that $[y, y + k] \subseteq [x_j - h_j, x_j]$ for some j, and

$$\frac{F(y + k) - F(y)}{k} > r. \tag{17.36}$$

Again by Lemma 17.3, there exists a finite sub-collection of disjoint intervals $[y_i, y_i + k_i]$, $i = 1, 2, \ldots, M$, such that

$$[y_i, y_i + k_i] \subseteq [x_j - h_j, x_j] \quad \text{for some } j, \tag{17.37}$$

$$F(y_i + k_i) - F(y_i) > r k_i, \tag{17.38}$$

and

$$\sum_{i=1}^{M} k_i \geq m(B) - \delta \geq m(A) - 3\delta, \tag{17.39}$$

where we have used (17.33). Combining (17.38) and (17.39) we have

$$\sum_{i=1}^{M} [F(y_i + k_i) - F(y_i)] > r(m(A) - 3\delta). \tag{17.40}$$

(c) Contradiction. Now we use that F is increasing. By (17.37), the disjoint intervals $[y_i, y_i + k_i]$ sit inside the intervals $[x_j - h_j, x_j]$. For each fixed j, the sum of the increments of F over those $[y_i, y_i + k_i]$ lying inside $[x_j - h_j, x_j]$ is at most $F(x_j) - F(x_j - h_j)$. Summing over j and using (17.40) and (17.35),

$$r(m(A) - 3\delta) < \sum_{i=1}^{M} \{F(y_i + k_i) - F(y_i)\} \tag{17.41}$$

$$\leq \sum_{j=1}^{N} \{F(x_j) - F(x_j - h_j)\} < s(m(A) + \delta). \tag{17.42}$$

Since $\delta > 0$ is arbitrary, letting $\delta \to 0$ gives $r\, m(A) \leq s\, m(A)$. As $m(A) > 0$, this means $r \leq s$, contradicting $r > s$. Hence $m(A) = 0$.

Step 2.2: all Dini numbers are finite and equal almost everywhere. So far we have

shown that

$$D_1F(x) < \infty \text{ for a.e. } x, \text{ and} \tag{17.43}$$

$$D_1F(x) \leq D_4F(x) \text{ for a.e. } x. \tag{17.44}$$

Applying (17.44) to $G(x) = -F(-x)$ (which is also increasing in $[-b, -a]$), we have

$$\limsup_{h\to 0^+} \frac{G(x+h) - G(x)}{h} \leq \liminf_{h\to 0^+} \frac{G(x) - G(x-h)}{h} \text{ for a.e. } x. \tag{17.45}$$

After simplifying (and putting $y = -x$), we get

$$\limsup_{h\to 0^+} \frac{-F(-x-h) + F(-x)}{h} \leq \liminf_{h\to 0^+} \frac{-F(-x) + F(-x+h)}{h} \text{ for a.e. } x \tag{17.46}$$

$$\implies \limsup_{h\to 0^+} \frac{F(y) - F(y-h)}{h} \leq \liminf_{h\to 0^+} \frac{F(y+h) - F(y)}{h} \text{ for a.e. } y \tag{17.47}$$

$$\implies D_3F(y) \leq D_2F(y) \text{ for a.e. } y. \tag{17.48}$$

Since we automatically have $D_1F \geq D_2F$ and $D_3F \geq D_4F$ (since supremum always beats infimum), combining all these, we get

$$D_1F \leq D_4F \leq D_3F \leq D_2F \leq D_1F < \infty \text{ a.e.} \tag{17.49}$$

Therefore, all Dini numbers are equal a.e. and finite a.e. Hence, F is differentiable a.e.

Step 3: measurability and the integral bound. By Steps 1 and 2, F is differentiable a.e. in (a, b). Now consider the difference quotients

$$F_k(x) = \frac{F\left(x + \frac{1}{k}\right) - F(x)}{\frac{1}{k}} \tag{17.50}$$

for $x \in (a, b)$, where we define $F(y) = F(b)$ for $y \geq b$ (otherwise $x + \frac{1}{k}$ might fall outside the interval when x is close to b). Note that each F_k is non-negative (since F is increasing) and measurable (it is a difference of measurable functions), and that $F_k \to F'$ a.e. in (a, b) as $k \to \infty$. Hence F' is a pointwise limit of non-negative measurable functions, so it is measurable and non-negative. Finally, by Fatou's Lemma,

$$\int_a^b F'(y)\,\mathrm{d}y \leq \liminf_{k\to\infty} \int_a^b F_k(y)\,\mathrm{d}y \tag{17.51}$$

$$= \liminf_{k\to\infty} \frac{1}{1/k} \left[\int_a^b F\left(y + \frac{1}{k}\right)\mathrm{d}y - \int_a^b F(y)\,\mathrm{d}y\right]. \tag{17.52}$$

We substitute $u = y + \frac{1}{k}$ in the integral of $F\left(y + \frac{1}{k}\right)$, obtaining

$$\int_a^b F\left(y + \frac{1}{k}\right)\mathrm{d}y = \int_{a+\frac{1}{k}}^{b+\frac{1}{k}} F(u)\,\mathrm{d}u. \tag{17.53}$$

Subtracting $\int_a^b F$, we get

$$\int_a^b F'(y)\,\mathrm{d}y \le \liminf_{k\to\infty} \frac{1}{1/k}\left[\int_b^{b+\frac{1}{k}} F(y)\,\mathrm{d}y - \int_a^{a+\frac{1}{k}} F(y)\,\mathrm{d}y\right]. \tag{17.54}$$

For $y \ge b$, $F(y) = F(b)$. So $\int_b^{b+\frac{1}{k}} F(y)\,\mathrm{d}y$ is an integral of a constant function. Therefore,

$$\frac{1}{1/k}\int_b^{b+\frac{1}{k}} F(y)\,\mathrm{d}y = F(b). \tag{17.55}$$

Since F is increasing, in the interval $\left[a, a+\frac{1}{k}\right]$, $F(y)$ is at least $F(a)$. Therefore,

$$\limsup_{k\to\infty} \frac{1}{1/k}\int_a^{a+\frac{1}{k}} F(y)\,\mathrm{d}y \ge F(a). \tag{17.56}$$

So we have

$$\int_a^b F'(y)\,\mathrm{d}y \le \liminf_{k\to\infty} \frac{1}{1/k}\left[\int_b^{b+\frac{1}{k}} F(y)\,\mathrm{d}y - \int_a^{a+\frac{1}{k}} F(y)\,\mathrm{d}y\right] \tag{17.57}$$

$$= F(b) - \limsup_{k\to\infty} \frac{1}{1/k}\int_a^{a+\frac{1}{k}} F(y)\,\mathrm{d}y \tag{17.58}$$

$$\le F(b) - F(a). \tag{17.59}$$

This concludes the proof of the theorem. ■

Remark 17.3. We do not necessarily have equality in the last step. If F happens to be continuous at a, the second average tends to $F(a)$, but the inequality $\int_a^b F' \le F(b) - F(a)$ may still be strict.

Corollary 17.4

If F is a function of bounded variation in $[a,b]$, then F is differentiable a.e. in (a,b) and $F' \in L^1([a,b])$.

Proof. By Proposition 17.1, $F = F_1 - F_2$ with F_1, F_2 bounded and increasing. By Theorem 17.2, each F_i is differentiable a.e. with F_i' non-negative, measurable and integrable by (17.14). Hence $F' = F_1' - F_2'$ a.e. and $F' \in L^1([a,b])$. ■

Functions of bounded variation are differentiable almost everywhere, but they do not necessarily satisfy the Fundamental theorem of calculus; we only have the inequality (17.14).

§18. 4 November 2026

Remark 18.1 (Cantor–Lebesgue function)**.** In general, for bounded and monotone functions, one cannot expect more than what was proved in Theorem 17.2. The Cantor–Lebesgue function $F : [0,1] \to [0,1]$ is an example of an increasing function with $F(0) = 0$, $F(1) = 1$ and such that $F'(x) = 0$ a.e. in $[0,1]$; in particular

$$\int_0^1 F'(y)\,\mathrm{d}y = 0 < 1 = F(1) - F(0), \tag{18.1}$$

so the inequality (17.14) can be strict.

In order to describe this function, consider the classical Cantor set $\mathcal{C}$, formed by removing the open middle third intervals, i.e.

$$\mathcal{C} = \bigcap_{k=1}^{\infty} C_k, \tag{18.2}$$

where C_k is the union of the 2^k closed intervals of the k-th step of the construction, i.e.

$$C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right], \qquad C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right], \tag{18.3}$$

and so on. Let us consider a sequence of increasing functions $\{F_k\}_{k\geq 1}$ defined as follows:

$$F_1(x) = \begin{cases} 0, & \text{if } x = 0, \\ \frac{1}{2}, & \text{if } \frac{1}{3} \leq x \leq \frac{2}{3}, \\ 1, & \text{if } x = 1, \end{cases} \tag{18.4}$$

and F_1 linear in C_1. Similarly,

$$F_2(x) = \begin{cases} 0, & \text{if } x = 0, \\ \frac{1}{4}, & \text{if } \frac{1}{9} \leq x \leq \frac{2}{9}, \\ \frac{1}{2}, & \text{if } \frac{1}{3} \leq x \leq \frac{2}{3}, \\ \frac{3}{4}, & \text{if } \frac{7}{9} \leq x \leq \frac{8}{9}, \\ 1, & \text{if } x = 1, \end{cases} \tag{18.5}$$

and F_2 linear in C_2, and so on. The sequence $\{F_k\}_{k\geq 1}$ then verifies

$$|F_{k+1}(x) - F_k(x)| \leq 2^{-k-1}, \tag{18.6}$$

and hence $\{F_k\}_{k\geq 1}$ is uniformly convergent to a limiting function F that is continuous, increasing, with $F(0) = 0$ and $F(1) = 1$, and constant in each interval of the complement of $\mathcal{C}$. Since $m(\mathcal{C}) = 0$, we have $F'(x) = 0$ a.e. in $(0,1)$.

§18.i. Absolutely continuous functions

Definition 18.1 (Absolute continuity). We say that a function $F : [a, b] \to \mathbb{R}$ is **absolutely continuous** if, given $\varepsilon > 0$, there exists $\delta > 0$ such that for any countable collection (finite or not) of disjoint intervals $[a_i, b_i] \subseteq [a, b]$ we have

$$\sum_i (b_i - a_i) < \delta \quad \Longrightarrow \quad \sum_i |F(b_i) - F(a_i)| < \varepsilon. \tag{18.7}$$

A basic example of an absolutely continuous function is given by the integral of a function $g \in L^1[a, b]$, i.e.

$$G(x) = \int_a^x g(y)\,\mathrm{d}y. \tag{18.8}$$

The absolute continuity of G follows from Exercise 10.8. In this case, Lebesgue differentiation theorem tells us that $G'(x) = g(x)$ a.e. in (a, b). The objective of this section is to prove the converse: that any absolutely continuous function is given by the integral of an L^1-function (hence proving that the Fundamental theorem of calculus holds if and only if F is absolutely continuous).

Proposition 18.1

If $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then it has bounded variation.

Proof. Let $\varepsilon = 1$, and δ be given by absolute continuity. Let $N = \left\lceil \frac{b-a}{\delta} \right\rceil + 1$. Then we do a partition of $[a, b]$ as

$$x_i = a + i\,\frac{b-a}{N}, \qquad \text{for } i = 0, 1, \ldots, N. \tag{18.9}$$

Then $|x_i - x_{i-1}| = \frac{b-a}{N} < \delta$. Therefore, any partition of $[x_{i-1}, x_i]$ will have sum of the length of the intervals $< \delta$. Therefore, by absolute continuity,

$$\mathrm{Var}_F[x_{i-1}, x_i] \leq 1. \tag{18.10}$$

Then by Remark 17.1(4), we have

$$\mathrm{Var}_F[a, b] = \sum_{i=1}^N \mathrm{Var}_F[x_{i-1}, x_i] \leq \sum_{i=1}^N 1 = N. \tag{18.11}$$

Hence, $\mathrm{Var}_F[a, b] < \infty$, i.e. F has bounded variation. ∎

In particular, if $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then F' is defined a.e. in (a, b) (by Corollary 17.4).

§18.ii. Fundamental theorem of calculus for AC functions

Theorem 18.2

Let $F : [a, b] \to \mathbb{R}$ be an absolutely continuous function such that $F' = 0$ a.e. in (a, b). Then

F is constant.

Proof. It suffices to prove that $F(b) = F(a)$, since the same argument can be applied to any subinterval. Let $E \subseteq (a,b)$ be the subset where $F' = 0$, and hence $m(E) = b - a$.

Given $\varepsilon > 0$ and $x \in E$, there exists a sequence of arbitrarily small $h > 0$ such that $[x, x+h] \subseteq (a,b)$ and $|F(x+h) - F(x)| < \varepsilon h$. Let $\delta > 0$ be the number corresponding to this $\varepsilon > 0$ in the definition of absolute continuity.

The intervals $[x, x+h]$ form a Vitali covering of E, so by Vitali covering, second version (applied with $\frac{\delta}{2}$) there exists a finite sub-collection of disjoint intervals $[x_i, x_i + h_i]$ with $i = 1, 2, \ldots, N$ such that

$$|F(x_i + h_i) - F(x_i)| < \varepsilon h_i \tag{18.12}$$

and

$$\sum_{i=1}^{N} h_i \geq (b-a) - \frac{\delta}{2}. \tag{18.13}$$

Since $[x_i, x_i + h_i] \subseteq (a,b)$, we have that $\sum_{i=1}^{N} h_i \leq b - a$. Therefore, summing up (18.12), we have

$$\sum_{i=1}^{N} |F(x_i + h_i) - F(x_i)| < \varepsilon \sum_{i=1}^{N} h_i \leq \varepsilon (b-a). \tag{18.14}$$

On the other hand, using (18.13), we know that the total length of the complementary intervals (the finitely many intervals into which $[a,b]$ is split by the $[x_i, x_i + h_i]$) is less than or equal to $\frac{\delta}{2} < \delta$, and from the definition of absolute continuity, the sum of the increments of F in these intervals is less than or equal to ε. Since $F(b) - F(a)$ is the sum of the increments of F over the intervals $[x_i, x_i + h_i]$ and over the complementary intervals, from the triangle inequality we then have

$$|F(b) - F(a)| < \varepsilon (b-a) + \varepsilon, \tag{18.15}$$

and since $\varepsilon > 0$ is arbitrary, we conclude that $F(b) = F(a)$. ∎

Theorem 18.3 (Fundamental theorem of calculus for the Lebesgue integral)

A function $F : [a,b] \to \mathbb{R}$ is absolutely continuous in $[a,b]$ if and only if F' exists a.e. in (a,b), $F' \in L^1([a,b])$ and

$$F(x) - F(a) = \int_a^x F'(y)\,\mathrm{d}y, \tag{18.16}$$

for any $a \leq x \leq b$.

Proof. We have already seen that any function of the form $G(x) = \int_a^x g(y)\,\mathrm{d}y$, where $g \in L^1([a,b])$, is absolutely continuous, and from Lebesgue differentiation theorem we have $G' = g$ a.e. in (a,b). This gives the "if" direction.

On the other hand, suppose that F is absolutely continuous and define

$$G(x) = \int_a^x F'(y)\,\mathrm{d}y. \tag{18.17}$$

We have already seen that $F' \in L^1([a,b])$ (since F has bounded variation, see Corollary 17.4), and hence G is absolutely continuous and, by Lebesgue differentiation theorem, we have $G' = F'$ a.e. in (a,b).

It then follows that $F - G$ is absolutely continuous and $(F-G)' = 0$ a.e. in (a,b). From Theorem 18.2 we conclude that $F - G$ is constant, and since $G(a) = 0$ we arrive at

$$F(x) - G(x) = F(a). \tag{18.18}$$

■

§18.iii. Exercises

Exercise 18.1. Let $f : [0,1] \to \mathbb{R}$ be an absolutely continuous function such that

$$\int_0^1 f(x)\,\mathrm{d}x = 0. \tag{18.19}$$

Show that

$$\sup_{0\le y\le 1}\left|\int_0^1 (x-y)\,f'(x)\,\mathrm{d}x\right| \le \sup_{0\le x\le 1}|f(x)|. \tag{18.20}$$

Solution. We can apply integration by parts to evaluate $\int_0^1 (x-y)\,f'(x)\,\mathrm{d}x$, where $u = x-y$ and $\mathrm{d}v = f'(x)\,\mathrm{d}x$, so that $v = f(x)$. Then

$$\int_0^1 (x-y)\,f'(x)\,\mathrm{d}x = \int_{x=0}^{x=1} u\,\mathrm{d}v = \Big[uv\Big]_{x=0}^{x=1} - \int_{x=0}^{x=1} v\,\mathrm{d}u \tag{18.21}$$

$$= \Big[(x-y)f(x)\Big]_{x=0}^{x=1} - \int_0^1 f(x)\,\mathrm{d}x \tag{18.22}$$

$$= (1-y)\,f(1) + yf(0). \tag{18.23}$$

Let $M = \sup_{0\le x\le 1}|f(x)|$. Then

$$\left|\int_0^1 (x-y)\,f'(x)\,\mathrm{d}x\right| \le (1-y)\,|f(1)| + y\,|f(0)| \le (1-y)\,M + yM = M. \tag{18.24}$$

■

Exercise 18.2. Let $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function with $f' \in L^1(\mathbb{R})$. Show that for all $h \in \mathbb{R}$ we have

$$\int_{-\infty}^{\infty} |f(x+h) - f(x)|\,\mathrm{d}x \le C\,|h|, \tag{18.25}$$

where C is a constant independent of h.

Solution. First, WLOG, we can assume $h > 0$, because if we had $-h$ in place of h, (with $h > 0$),

$$\int_{-\infty}^{\infty} |f(x-h) - f(x)| \, \mathrm{d}x = \int_{-\infty}^{\infty} |f(x) - f(x-h)| \, \mathrm{d}x \tag{18.26}$$

$$= \int_{-\infty}^{\infty} |f(y+h) - f(y)| \, \mathrm{d}y. \tag{18.27}$$

If we prove that it's $\leq C|h|$ for $h \geq 0$, then we are done! So from now on, we'll assume $h > 0$.

By Fundamental theorem of calculus for the Lebesgue integral,

$$f(x+h) - f(x) = \int_x^{x+h} f'(u) \, \mathrm{d}u = \int_0^h f'(x+t) \, \mathrm{d}t, \tag{18.28}$$

where $u = x + t$. Then we have

$$|f(x+h) - f(x)| = \left| \int_0^h f'(x+t) \, \mathrm{d}t \right| \leq \int_0^h \left| f'(x+t) \right| \, \mathrm{d}t. \tag{18.29}$$

Therefore,

$$\text{LHS} = \int_{-\infty}^{\infty} |f(x+h) - f(x)| \, \mathrm{d}x \leq \int_{-\infty}^{\infty} \int_0^h \left| f'(x+t) \right| \, \mathrm{d}t \, \mathrm{d}x. \tag{18.30}$$

By Tonelli's theorem, we can switch the order of integrations:

$$\text{LHS} \leq \int_0^h \left[\int_{-\infty}^{\infty} \left| f'(x+t) \right| \, \mathrm{d}x \right] \, \mathrm{d}t \tag{18.31}$$

$$= \int_0^h \left[\int_{-\infty}^{\infty} \left| f'(u) \right| \, \mathrm{d}u \right] \, \mathrm{d}t \tag{18.32}$$

$$\leq \left\| f' \right\|_{L^1(\mathbb{R})} \int_0^h \mathrm{d}t = \left\| f' \right\|_{L^1(\mathbb{R})} h. \tag{18.33}$$

(Here we have done a chanmge of variables $u = x + t$, so that $\mathrm{d}x = \mathrm{d}u$.) Therefore, (18.25) holds, with $C = \|f'\|_{L^1(\mathbb{R})}$ is the desired constant that doesn't depend on h. ∎

Exercise 18.3. Let $f : [0,1] \to \mathbb{R}$ be an absolutely continuous function such that $f' \in L^2([0,1])$. Show that for each $\varepsilon > 0$ there exists $\delta > 0$ such that: if $0 \leq y < x \leq 1$ with $|x - y| < \delta$ then

$$\frac{|f(x) - f(y)|^2}{|x - y|} < \varepsilon. \tag{18.34}$$

Solution. Since f is absolutely continuous,

$$f(x) - f(y) = \int_y^x f'(t) \, \mathrm{d}t. \tag{18.35}$$

Then we'll apply Cauchy–Schwarz inequality (or Hölder's inequality).

$$|f(x)-f(y)|^2 = \left|\int_y^x f'(t)\,\mathrm{d}t\right|^2 \tag{18.36}$$

$$\leq \left[\int_y^x 1^2\,\mathrm{d}t\right]\left[\int_y^x \left|f'(t)\right|^2\,\mathrm{d}t\right] \tag{18.37}$$

$$= |x-y|\int_y^x \left|f'(t)\right|^2\,\mathrm{d}t. \tag{18.38}$$

We are given that $|f'|^2 \in L^1([0,1])$. Therefore, $|f'|^2 \in L^1([0,1])$. By absolute continuity of integral of an L^1 function (Exercise 10.8), for every ε, there exists $\delta > 0$ such that

$$\left|\int_E |f'|^2\,\mathrm{d}m\right| \leq \varepsilon \text{ whenever } m(E) < \delta. \tag{18.39}$$

Since the measure of the interval $[y,x] = x-y < \delta$, we have

$$\int_y^x \left|f'(t)\right|^2\,\mathrm{d}t \leq \varepsilon. \tag{18.40}$$

Therefore,

$$|f(x)-f(y)|^2 \leq |x-y|\,\varepsilon, \tag{18.41}$$

and we are done! ■

Exercise 18.4. Let φ be a Lebesgue measurable function in $\mathbb{R}^d$ that verifies the following property: for each d-dimensional rectangle Q we have

$$\left|\int_Q \varphi(x)\,\mathrm{d}x\right| \leq \frac{M\,m(Q)}{1+m(Q)}, \tag{18.42}$$

where M is a constant. Show that for each $f \in L^1\left(\mathbb{R}^d\right)$ we have

$$\lim_{k\to\infty}\int_{\mathbb{R}^d} \varphi(kx)\,f(x)\,\mathrm{d}x = 0. \tag{18.43}$$

Solution. Since integral and limits are linear, (18.43) is linear in f. If $f = \chi_Q$ is the characteristic function of a rectangle. Then

$$I := \int_{\mathbb{R}^d} \varphi(kx)\,\chi_Q(x)\,\mathrm{d}x = \int_Q \varphi(kx)\,\mathrm{d}x = \frac{1}{k^d}\int_{kQ}\varphi(u)\,\mathrm{d}u. \tag{18.44}$$

Then the absolute value of this is

$$|I| = \frac{1}{k^d}\left|\int_{kQ}\varphi(u)\,\mathrm{d}u\right| \leq \frac{1}{k^d}\frac{M\,m(kQ)}{1+m(kQ)} = \frac{1}{k^d}\frac{Mk^d\,m(Q)}{1+k^d\,m(Q)} = \frac{M\,m(Q)}{1+k^d\,m(Q)}. \tag{18.45}$$

As $k \to \infty$, this approaches 0. Hence,

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} \varphi(kx)\, \chi_Q(x)\, \mathrm{d}x = 0. \tag{18.46}$$

Hence, (18.43) holds for χ_Q, and by linearity, it holds for every step-function ψ. Now let $f \in L^1$ be arbitrary, and fix $\varepsilon > 0$. Since the class of step functions is dense in $L^1\left(\mathbb{R}^d\right)$, we can pick a step function ψ such that

$$\|f - \psi\|_1 \leq \frac{\varepsilon}{2M}. \tag{18.47}$$

First, we'll show that $|\varphi| \leq M$ a.e. For that, we'll apply Lebesgue differentiation theorem along with Remark 16.7. Let Q_ε be a cube centered at x with diameter ε. By Lebesgue differentiation theorem (it works for the collection of cubes as well), for a.e. x,

$$\varphi(x) = \lim_{\varepsilon\to 0} \frac{1}{m(Q_\varepsilon)} \int_{Q_\varepsilon} \varphi(y)\, \mathrm{d}y. \tag{18.48}$$

Then

$$|\varphi(x)| = \lim_{\varepsilon\to 0} \frac{1}{m(Q_\varepsilon)} \left| \int_{Q_\varepsilon} \varphi(y)\, \mathrm{d}y \right| \tag{18.49}$$

$$\leq \lim_{\varepsilon\to 0} \frac{1}{m(Q_\varepsilon)} \frac{M\, m(Q_\varepsilon)}{1 + m(Q_\varepsilon)} \tag{18.50}$$

$$= M \lim_{\varepsilon\to 0} \frac{1}{1 + m(Q_\varepsilon)} = M. \tag{18.51}$$

Now,

$$I := \int_{\mathbb{R}^d} \varphi(kx)\, f(x)\, \mathrm{d}x = \int_{\mathbb{R}^d} \varphi(kx)\, \psi(x)\, \mathrm{d}x + \int_{\mathbb{R}^d} \varphi(kx)\, [f(x) - \psi(x)]\, \mathrm{d}x. \tag{18.52}$$

Since the first summand approaches 0, there exists N such that for $k \geq N$,

$$\left| \int_{\mathbb{R}^d} \varphi(kx)\, \psi(x)\, \mathrm{d}x \right| \leq \frac{\varepsilon}{2}. \tag{18.53}$$

Then we have

$$|I| \leq \left| \int_{\mathbb{R}^d} \varphi(kx)\, \psi(x)\, \mathrm{d}x \right| + \int_{\mathbb{R}^d} |\varphi(kx)|\, |f(x) - \psi(x)|\, \mathrm{d}x \tag{18.54}$$

$$\leq \frac{\varepsilon}{2} + M \int_{\mathbb{R}^d} |f(x) - \psi(x)|\, \mathrm{d}x \tag{18.55}$$

$$\leq \frac{\varepsilon}{2} + M\, \frac{\varepsilon}{2M} = \varepsilon. \tag{18.56}$$

Therefore, $\int_{\mathbb{R}^d} \varphi(kx)\, f(x)\, \mathrm{d}x \to 0$ as $k \to \infty$. ■

Exercise 18.5. Let $f_n : \mathbb{R} \to [0, \infty)$ be non-decreasing functions for each $n \in \mathbb{N}$. Suppose that for all $x \in \mathbb{R}$ we have

$$f(x) = \sum_{n=1}^{\infty} f_n(x) < \infty. \tag{18.57}$$

Prove that

$$f'(x) = \sum_{n=1}^{\infty} f_n'(x) \quad \text{a.e.} \tag{18.58}$$

Solution. Since each f_n is non-decreasing,

$$\frac{f_n(x+h) - f_n(x)}{h} \geq 0. \tag{18.59}$$

Since $f = \sum_{n=1}^{\infty} f_n$, we have

$$\frac{f(x+h) - f(x)}{h} = \sum_{n=1}^{\infty} \frac{f_n(x+h) - f_n(x)}{h}. \tag{18.60}$$

Applying Fatou's Lemma for series (series can been seen as integral, where the measure space is $\mathbb{N}$ with respect to the counting measure),

$$\sum_{n=1}^{\infty} f_n'(x) = \sum_{n=1}^{\infty} \lim_{h\to 0} \frac{f_n(x+h) - f_n(x)}{h} \tag{18.61}$$

$$\leq \liminf_{h\to 0} \sum_{n=1}^{\infty} \frac{f_n(x+h) - f_n(x)}{h} \tag{18.62}$$

$$= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = f'(x). \tag{18.63}$$

Let $R_N = \sum_{n=N+1}^{\infty} f_n$. Then $f = \sum_{n=1}^{N} f_n + R_N$, and differentiation is linear, so

$$f' = \sum_{n=1}^{N} f_n' + R_N'. \tag{18.64}$$

Integrating over $[a, b]$, we have

$$\int_a^b f'(x)\,\mathrm{d}x = \sum_{n=1}^{N} \int_a^b f_n'(x)\,\mathrm{d}x + \int_a^b R_N'(x)\,\mathrm{d}x \tag{18.65}$$

$$\leq \sum_{n=1}^{N} \int_a^b f_n'(x)\,\mathrm{d}x + [R_N(b) - R_N(a)]. \tag{18.66}$$

Taking $N \to \infty$, $R_N \to 0$. So we have

$$\int_a^b f'(x)\,\mathrm{d}x \leq \sum_{n=1}^{\infty} \int_a^b f_n'(x)\,\mathrm{d}x = \int_a^b \sum_{n=1}^{\infty} f_n'(x)\,\mathrm{d}x. \tag{18.67}$$

Here we used Proposition 2.8 to pass the sum inside the integral, since $f_n' \geq 0$. Therefore,

$$\int_a^b \left[f'(x) - \sum_{n=1}^{\infty} f_n'(x) \right] \mathrm{d}x \leq 0. \tag{18.68}$$

Since the integrand is ≥ 0 a.e. by (18.63), we can conclude that

$$\int_a^b \left[f'(x) - \sum_{n=1}^{\infty} f_n'(x) \right] \mathrm{d}x = 0. \tag{18.69}$$

This gives

$$f'(x) = \sum_{n=1}^{\infty} f_n'(x) \text{ for a.e. } x \in [a,b]. \tag{18.70}$$

Since we can write $\mathbb{R}$ as a countable union of closed intervals, say $\mathbb{R} = \bigcup_n [-n, n]$, this holds for almost every $x \in \mathbb{R}$. ■

§19. 5 November 2026

§19.i. The convolution operator

Definition 19.1 (Convolution)**.** Let f and g be Lebesgue measurable functions on $\mathbb{R}^d$. The **convolution** of f and g is the function $f * g$ defined by

$$f * g\,(x) := \int_{\mathbb{R}^d} f\,(x-y)\,g\,(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f\,(y)\,g\,(x-y)\,\mathrm{d}y. \tag{19.1}$$

The two expressions agree after the change of variables $z = x - y$. So it does not matter whether we put x in the first or in the second function.

Remark 19.1 (Interpretation)**.** One should see the convolution as an **averaging** operator. In the first expression in (19.1), $f * g\,(x)$ is an average of f around the point x: the value of f at the point $x - y$ is weighted by $g\,(y)$. By symmetry, it is also an average of g around x with weights coming from f. So $f * g$ is a kind of weighted average.

If f and g are non-negative, then $f * g$ is a measurable function (possibly taking the value $+\infty$ at some points), so everything is well defined. If f and g can take both signs, we need some control on their growth/decay to make sure that the integral is a well-defined number; we shall see how to guarantee this below (at least almost everywhere), using integrability conditions on f and g.

For $w \in \mathbb{R}^d$ we denote the translation of a function by $\tau_w f\,(x) := f\,(x-w)$.

Proposition 19.1 (Basic properties)

Assuming that the integrals below are well defined, we have:

(i) the convolution operator is bilinear and symmetric: $f * g = g * f$;

(ii) it is associative: $(f * g) * h = f * (g * h)$;

(iii) for $w \in \mathbb{R}^d$ we have $\tau_w\,(f * g) = (\tau_w f) * g = f * (\tau_w g)$;

(iv) if A is the closure of the set $\{x + y : x \in \operatorname{supp} f,\ y \in \operatorname{supp} g\}$, then $\operatorname{supp}\,(f * g) \subseteq A$. In particular, if f and g have compact support, so does $f * g$.

Proof. (i) Bilinearity follows from the linearity of the integral: for instance, replacing g by $g_1 + g_2$ in the definition, the integral splits into two. Symmetry is exactly the equality of the two expressions in (19.1).

(ii) We expand both sides. On one hand,

$$((f * g) * h)\,(x) = \int_{\mathbb{R}^d} (f * g)\,(x-z)\,h\,(z)\,\mathrm{d}z = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f\,(x-z-y)\,g\,(y)\,h\,(z)\,\mathrm{d}y\,\mathrm{d}z. \tag{19.2}$$

On the other hand,

$$(f*(g*h))(x) = \int_{\mathbb{R}^d} f(x-u)\,(g*h)(u)\,\mathrm{d}u = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x-u)\,g(u-z)\,h(z)\,\mathrm{d}z\,\mathrm{d}u, \quad (19.3)$$

and for fixed z, the change of variables $u = y + z$ in the inner integral $\mathrm{d}u$ gives the same double integral as before, after exchanging the order of integration (Tonelli's theorem/Fubini's theorem).

(iii) We have

$$\tau_w (f*g)(x) = \int_{\mathbb{R}^d} f(x-w-y)\,g(y)\,\mathrm{d}y = ((\tau_w f)*g)(x), \quad (19.4)$$

and, using the second expression in (19.1),

$$\tau_w (f*g)(x) = \int_{\mathbb{R}^d} f(y)\,g(x-w-y)\,\mathrm{d}y = (f*(\tau_w g))(x). \quad (19.5)$$

(iv) We can only have $f(x-y)\,g(y) \neq 0$ if $y \in \operatorname{supp} g$ and $x - y \in \operatorname{supp} f$, in which case $x = (x-y) + y$ belongs to the set whose closure is A. Hence, if $x \notin A$, the integrand in (19.1) vanishes (almost everywhere in y) and $f*g(x) = 0$. So $f*g$ vanishes on the open set $\mathbb{R}^d \setminus A$, and therefore $\operatorname{supp}(f*g) \subseteq A$. If $\operatorname{supp} f$ and $\operatorname{supp} g$ are compact, then the set of sums is compact (it is the image of the compact set $\operatorname{supp} f \times \operatorname{supp} g$ under the continuous map $(x,y) \mapsto x+y$), so A is compact. ■

§19.ii. The regularizing effect

One of the most important features of the convolution is that it is a *regularizing* operator: as a broad guideline, the convolution of two functions inherits at least the regularity of the most regular of the two functions.

Proposition 19.2 (Regularization)

Let $f \in L^1(\mathbb{R}^d)$ be a possibly very irregular function, and let $g \in C^1_c(\mathbb{R}^d)$ (a function with compact support and a continuous derivative). Then $f*g \in C^1(\mathbb{R}^d)$ and

$$\partial_j (f*g) = f*(\partial_j g), \qquad j = 1,\dots,d. \quad (19.6)$$

Proof. Since g is bounded, $\int |f(y)|\,|g(x-y)|\,\mathrm{d}y \leq \|g\|_{L^\infty}\|f\|_{L^1} < \infty$, so $f*g(x)$ is defined for *every* x.

We want to show that the derivative exists, so we look at the difference quotient in the direction e_j. Which of the two expressions in (19.1) should we use? The point x appears inside g in the second one, and g is the good function, so we put the responsibility on the good function and use the second expression: for $h \neq 0$,

$$\frac{f*g(x+he_j) - f*g(x)}{h} = \int_{\mathbb{R}^d} f(y)\,\frac{g(x+he_j-y) - g(x-y)}{h}\mathrm{d}y. \quad (19.7)$$

We want to let $h \to 0$ inside the integral. By the mean value theorem, the difference quotient of g is equal to $\partial_j g$ evaluated at some intermediate point, hence

$$\left| \frac{g(x + he_j - y) - g(x - y)}{h} \right| \leq \|\partial_j g\|_{L^\infty} < \infty, \tag{19.8}$$

where the supremum is finite because $\partial_j g$ is continuous with compact support. Therefore the integrand is dominated by $\|\partial_j g\|_{L^\infty} |f(y)|$, which belongs to L^1 and does not depend on h. For each fixed y, the difference quotient of g converges to $\partial_j g(x - y)$. By Dominated Convergence Theorem, the limit exists and

$$\partial_j (f * g)(x) = \int_{\mathbb{R}^d} f(y) \, \partial_j g(x - y) \, dy = f * (\partial_j g)(x). \tag{19.9}$$

Finally, the partial derivatives are continuous: $G := \partial_j g$ is continuous with compact support, hence uniformly continuous, and

$$|f * G(x + w) - f * G(x)| \leq \int_{\mathbb{R}^d} |f(y)| \, |G(x + w - y) - G(x - y)| \, dy \tag{19.10}$$

$$\leq \|f\|_{L^1} \sup_{z \in \mathbb{R}^d} |G(z + w) - G(z)| \to 0 \tag{19.11}$$

as $w \to 0$, uniformly in x. The same argument (with g in place of G) shows that $f * g$ is continuous. Hence $f * g \in C^1(\mathbb{R}^d)$. ■

Remark 19.2 (Moral)**.** We could pass one derivative from $f * g$ to the good function g. In general, if $f \in C_c^k(\mathbb{R}^d)$ and $g \in L^1_{\text{loc}}(\mathbb{R}^d)$, then $f * g \in C^k(\mathbb{R}^d)$ and

$$\partial^\alpha (f * g) = (\partial^\alpha f) * g \tag{19.12}$$

for every multi-index α with $|\alpha| \leq k$. In fact f need not be compactly supported. Even if $f \in C^k(\mathbb{R}^d)$ and $\partial^\alpha f \in L^\infty(\mathbb{R}^d)$ for each $|\alpha| \leq k$, and $g \in L^1_{\text{loc}}(\mathbb{R}^d)$, then $f * g \in C^k(\mathbb{R}^d)$. Because notice that we needed f compactly supported to deduce that $\|\partial^\alpha f\|_\infty < \infty$. This was a crucial step in applying Dominated Convergence Theorem.

Moreover, by symmetry we can put derivatives on *both* functions: for example, if $f \in C_c^2$ and $g \in C_c^3$, then $f * g$ has at least three derivatives (all put on f... or on g), and in fact five, since we can place two derivatives on one function and three on the other. So the convolution is a regularizing procedure: we take a possibly terrible function f, average it against a better function g, and we obtain a function which is at least as regular as the better one. Averaging smooths things out.

§19.iii. Convolution of L^p functions

We now look for results that control the integrability and regularity of $f * g$ when f and g lie in Lebesgue spaces. We first recall that translations are continuous in L^p.

Lemma 19.3 (Continuity of translations)

Let $f \in L^p\left(\mathbb{R}^d\right)$ with $1 \leq p < \infty$. Then

$$\lim_{|w| \to 0} \|\tau_w f - f\|_{L^p} = 0. \tag{19.13}$$

Proof. Exercise 10.5. ■

Proposition 19.4

Let $1 \leq p, q \leq \infty$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p\left(\mathbb{R}^d\right)$ and $g \in L^q\left(\mathbb{R}^d\right)$, then $f * g$ is well defined everywhere, bounded and uniformly continuous. Moreover, if $1 < p, q < \infty$, then $f * g \in C_0\left(\mathbb{R}^d\right)$ (the space of continuous functions that tend to zero as $|x| \to \infty$).

This is a first example of the regularizing effect: if f is in L^p, possibly a very wild function, and g is in L^q, possibly a very wild function, their convolution is a nice function. This is very similar to Exercise 10.16.

Proof. ***Boundedness.*** By Hölder's inequality in the variable y (the translated and reflected function $y \mapsto f(x-y)$ has the same L^p norm as f, and the integrals are in y, so the point x plays no role),

$$|f * g(x)| = \left| \int_{\mathbb{R}^d} f(x-y)\, g(y)\, \mathrm{d}y \right| \tag{19.14}$$

$$\leq \int_{\mathbb{R}^d} |f(x-y)|\, |g(y)|\, \mathrm{d}y \tag{19.15}$$

$$\leq \|f\|_{L^p} \|g\|_{L^q}. \tag{19.16}$$

So $f * g(x)$ is well defined for every x and $|f * g(x)| \leq \|f\|_{L^p} \|g\|_{L^q}$.

Uniform continuity. Assume first $p < \infty$. By Proposition 19.1 (iii) and bilinearity, $\tau_w(f * g) - f * g = (\tau_w f - f) * g$, and applying the bound just proved to the pair $(\tau_w f - f, g)$ we get

$$|(f * g)(x - w) - (f * g)(x)| = |(\tau_w f - f) * g(x)| \tag{19.17}$$

$$\leq \|\tau_w f - f\|_{L^p} \|g\|_{L^q} \to 0 \quad \text{as} \quad |w| \to 0 \tag{19.18}$$

by Continuity of translations. This is exactly uniform continuity of $f * g$ (note that the bound does not depend on the point). If $p = \infty$, then $q = 1 < \infty$, and we use the symmetry $f * g = g * f$ to change the roles of p and q in the argument above.

Decay at infinity. Assume $1 < p, q < \infty$ and let $\varepsilon > 0$. Since C_c is dense in L^p and in L^q, we can choose continuous functions with compact support $\tilde{f}$ and $\tilde{g}$ such that

$$\left\| f - \tilde{f} \right\|_{L^p} < \varepsilon \text{ and } \|g - \tilde{g}\|_{L^q} < \varepsilon \tag{19.19}$$

We add and subtract:

$$f * g = \left(f - \tilde{f}\right) * g + \tilde{f} * (g - \tilde{g}) + \tilde{f} * \tilde{g}. \tag{19.20}$$

By the boundedness estimate, for every x,

$$\left|\left(f - \tilde{f}\right) * g\,(x)\right| \leq \varepsilon \|g\|_{L^q}, \qquad \left|\tilde{f} * (g - \tilde{g})\,(x)\right| \leq \left\|\tilde{f}\right\|_{L^p} \varepsilon \leq (\|f\|_{L^p} + \varepsilon)\,\varepsilon. \tag{19.21}$$

For the last term, $\tilde{f}$ and $\tilde{g}$ have compact support, so by Proposition 19.1 (iv) the function $\tilde{f} * \tilde{g}$ has compact support: there is R with $\tilde{f} * \tilde{g}\,(x) = 0$ for $|x| > R$. Therefore, for $|x| > R$,

$$|f * g\,(x)| \leq \varepsilon \left(\|g\|_{L^q} + \|f\|_{L^p} + \varepsilon\right). \tag{19.22}$$

Since $\varepsilon > 0$ was arbitrary, $f * g\,(x) \to 0$ as $|x| \to \infty$. Together with continuity, $f * g \in C_0\left(\mathbb{R}^d\right)$. ■

Remark 19.3. The decay can fail at the endpoints: if $p = 1$ and $q = \infty$ we only get boundedness and uniform continuity, since in that case we cannot approximate $g \in L^\infty$ by compactly supported functions.

§19.iv. Young's inequality

Theorem 19.5 (Young's inequality)

Let $1 \leq p, q, r \leq \infty$ be such that

$$1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}. \tag{19.23}$$

If $f \in L^p\left(\mathbb{R}^d\right)$ and $g \in L^q\left(\mathbb{R}^d\right)$, then $f * g$ is finite almost everywhere in $\mathbb{R}^d$ and

$$\|f * g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q}. \tag{19.24}$$

This controls the integrability of $f * g$ in terms of the integrability of f and g. Note that Proposition 19.4 is the special case $r = \infty$, and that if f and g are allowed to oscillate (positive and negative values) we need some control of this type to know that $f * g$ is well defined, otherwise we may be dealing with a divergent integral.

Remark 19.4 (Why the relation (19.23)?)**.** This is the same reason as Remark 9.7. A global inequality of the form

$$\|f * g\|_{L^r} \leq C \|f\|_{L^p} \|g\|_{L^q}, \tag{19.25}$$

where C is a universal constant that doesn't depend on f, g, has to respect the symmetries of $\mathbb{R}^d$. Let $\lambda > 0$ and replace f by $f_\lambda(x) = f(\lambda x)$ and g by $g_\lambda(x) = g(\lambda x)$. From Remark 9.7 (9.19), we have

$$\|f_\lambda\|_{L^p} = \lambda^{-\frac{d}{p}} \|f\|_{L^p} \tag{19.26}$$

and

$$f_\lambda * g_\lambda(x) = \int_{\mathbb{R}^d} f(\lambda x - \lambda y)\, g(\lambda y)\, \mathrm{d}y = \lambda^{-d} (f * g)(\lambda x), \tag{19.27}$$

so we have

$$\|f_\lambda * g_\lambda\|_{L^r} = \lambda^{-d} \|(f*g)_\lambda\|_{L^r} = \lambda^{-d-\frac{d}{r}} \|f*g\|_{L^r}\,. \tag{19.28}$$

So (19.25) applied to f_λ, g_λ reads

$$\lambda^{-d-\frac{d}{r}} \|f*g\|_{L^r} \le C\lambda^{-\frac{d}{p}-\frac{d}{q}} \|f\|_{L^p} \|g\|_{L^q} \tag{19.29}$$

for all $\lambda > 0$, which can only hold (for a fixed pair f, g with $f * g \neq 0$) if the powers of λ coincide (again, same argument as Remark 9.7), i.e.

$$1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}. \tag{19.30}$$

This is the relation that keeps the powers on both sides compatible.

Proof of Young's inequality. Since $|f*g| \le |f| * |g|$ wherever the right-hand side is finite, and the norms of $|f|$ and f coincide, we may assume without loss of generality that f and g are non-negative. In this case $f * g$ is a well-defined measurable function with values in $[0, \infty]$, and once we prove (19.24), the right-hand side is finite, so $f * g < \infty$ almost everywhere.

The case $r = \infty$**.** Then $\frac{1}{p} + \frac{1}{q} = 1$, and (19.24) is exactly Hölder's inequality (this is the bound proved in Proposition 19.4 (19.16)).

The case $r < \infty$**.** First, observe what (19.23) tells us. Since $\frac{1}{q} \le 1$, we have

$$\frac{1}{p} = 1 + \frac{1}{r} - \frac{1}{q} \ge \frac{1}{r}, \tag{19.31}$$

and similarly $\frac{1}{q} \ge \frac{1}{r}$. In particular $p, q < \infty$ (as $\frac{1}{r} > 0$), and we can define $p_1, q_1 \in (1, \infty]$ by

$$\frac{1}{p_1} = \frac{1}{p} - \frac{1}{r}, \qquad \frac{1}{q_1} = \frac{1}{q} - \frac{1}{r}. \tag{19.32}$$

Then, by (19.23),

$$\frac{1}{r} + \frac{1}{p_1} + \frac{1}{q_1} = \frac{1}{p} + \frac{1}{q} - \frac{1}{r} = 1. \tag{19.33}$$

Now we break the integrand into three pieces, writing $f * g\,(x)$ in the form

$$f*g\,(x) = \int_{\mathbb{R}^d} \left[f\,(y)^{\frac{p}{r}}\, g\,(x-y)^{\frac{q}{r}} \right] f\,(y)^{p\left(\frac{1}{p}-\frac{1}{r}\right)}\, g\,(x-y)^{q\left(\frac{1}{q}-\frac{1}{r}\right)}\, \mathrm{d}y. \tag{19.34}$$

This is the same integral, since the exponent of $f\,(y)$ is $\frac{p}{r} + 1 - \frac{p}{r} = 1$, and likewise for $g\,(x-y)$. (When $p = r$, respectively $q = r$, the second, respectively third, function is identically 1.) Using

$$p\left(\frac{1}{p} - \frac{1}{r}\right) = \frac{p}{p_1} \text{ and } q\left(\frac{1}{q} - \frac{1}{r}\right) = \frac{q}{q_1}, \tag{19.35}$$

we apply Hölder's inequality to the three functions (with the three exponents r, p_1, q_1 satisfying

(19.33); this is Hölder applied twice):

$$f*g(x) \le \left(\int_{\mathbb{R}^d} f(y)^p\, g(x-y)^q \,\mathrm{d}y\right)^{\frac{1}{r}} \left(\int_{\mathbb{R}^d} f(y)^p \,\mathrm{d}y\right)^{\frac{1}{p_1}} \left(\int_{\mathbb{R}^d} g(x-y)^q \,\mathrm{d}y\right)^{\frac{1}{q_1}}. \tag{19.36}$$

(Indeed, $\left(f^{\frac{p}{p_1}}\right)^{p_1} = f^p$ and $\left(g^{\frac{q}{q_1}}\right)^{q_1} = g^q$.) The last two integrals are $\|f\|_{L^p}^p$ and $\|g\|_{L^q}^q$ (the point x disappears after the change of variables), so we obtain the pointwise inequality

$$f*g(x) \le \|f\|_{L^p}^{\frac{p}{p_1}} \|g\|_{L^q}^{\frac{q}{q_1}} \left(\int_{\mathbb{R}^d} f(y)^p\, g(x-y)^q \,\mathrm{d}y\right)^{\frac{1}{r}}. \tag{19.37}$$

Now we raise this to the power r and integrate in x. Using Tonelli's theorem, integrating first in x (which gives $\|g\|_{L^q}^q$) and then in y (which gives $\|f\|_{L^p}^p$),

$$\int_{\mathbb{R}^d} (f*g(x))^r \,\mathrm{d}x \le \|f\|_{L^p}^{\frac{rp}{p_1}} \|g\|_{L^q}^{\frac{rq}{q_1}} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(y)^p\, g(x-y)^q \,\mathrm{d}y\,\mathrm{d}x \tag{19.38}$$

$$= \|f\|_{L^p}^{\frac{rp}{p_1}+p} \|g\|_{L^q}^{\frac{rq}{q_1}+q}. \tag{19.39}$$

Taking the r-th root,

$$\|f*g\|_{L^r} \le \|f\|_{L^p}^{\frac{p}{p_1}+\frac{p}{r}} \|g\|_{L^q}^{\frac{q}{q_1}+\frac{q}{r}} = \|f\|_{L^p} \|g\|_{L^q}, \tag{19.40}$$

because

$$\frac{p}{p_1} + \frac{p}{r} = p\left(\frac{1}{p_1}+\frac{1}{r}\right) = p\cdot\frac{1}{p} = 1; \tag{19.41}$$

and in the same way $\frac{q}{q_1} + \frac{q}{r} = 1$. ∎

Remark 19.5 (Other proofs, and the sharp constant)**.**

1. The proof above is a "somewhat tricky" use of Hölder's inequality: one has to play around a bit to find the three-way splitting. A more conceptual proof goes through *interpolation*: we have already proved (19.24) for the endpoint $(p,q,r) = (p,p',\infty)$ (Proposition 19.4), and it is easy to prove it for $(p,q,r) = (1,1,1)$, since by Tonelli's theorem and the non-negativity of the functions

$$\int_{\mathbb{R}^d} f*g(x)\,\mathrm{d}x = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x-y)\,g(y)\,\mathrm{d}y\,\mathrm{d}x = \|f\|_{L^1}\|g\|_{L^1}. \tag{19.42}$$

   The interpolation theorem (which we do not see in this course) says that if a linear operator is bounded between certain pairs of L^p spaces, then it is bounded on the whole segment connecting these points (more precisely, in the space of points $\left(\frac{1}{p},\frac{1}{q}\right)$, on the convex hull), which then recovers (19.24) in the whole range.

2. The constant 1 in (19.24) is not sharp in general: the sharp constant was found by W. Beckner, it is attained for Gaussians, and it depends on p,q,r (and d). It is equal to 1 in the endpoint cases above. Finding sharp constants is a very interesting and much harder problem in analysis.

§19.v. Approximations of the identity

Let $\Phi : \mathbb{R}^d \to \mathbb{R}$ be a given function and let $\varepsilon > 0$. We write

$$\Phi_\varepsilon(x) := \frac{1}{\varepsilon^d}\Phi\left(\frac{x}{\varepsilon}\right). \tag{19.43}$$

Observe that if $\Phi \in L^1(\mathbb{R}^d)$, then, with the change of variables $x = \varepsilon y$ (so $\mathrm{d}x = \varepsilon^d \mathrm{d}y$),

$$\int_{\mathbb{R}^d} \Phi_\varepsilon(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \frac{1}{\varepsilon^d}\,\Phi\left(\frac{x}{\varepsilon}\right)\mathrm{d}x = \int_{\mathbb{R}^d} \Phi(y)\,\mathrm{d}y \tag{19.44}$$

for each $\varepsilon > 0$: the normalization ε^{-d} does not change the total weight. Picture, for Φ a bump supported in $[-1,1]$: Φ_ε is supported in $[-\varepsilon,\varepsilon]$, it is squeezed horizontally, and it is stretched vertically (its height is multiplied by ε^{-d}) so that the area remains the same.

Remark 19.6 (Heuristics: Dirac delta)**.** When $\varepsilon \to 0$, the mass of Φ_ε becomes more and more concentrated near the origin, in such a way that, heuristically, Φ_ε converges to a multiple of the Dirac delta (a *measure*, not a function: it gives mass 1 to a set if it contains the origin and 0 otherwise; integrating against δ_0 just evaluates the function at 0). The Dirac delta acts as the identity for the convolution operation: $f * \delta_0 = f$. Hence we expect that when $\int \Phi = 1$, we'll have $f * \Phi_\varepsilon \approx f$ for small ε. Another way to see this: if $\Phi \geq 0$ and $\int \Phi = 1$, then

$$f * \Phi_\varepsilon(x) = \int_{\mathbb{R}^d} f(x-y)\,\Phi_\varepsilon(y)\,\mathrm{d}y \tag{19.45}$$

is a *weighted average* of f around x, with total weight 1, and when $\varepsilon \to 0$ the weights concentrate near the point x. So it is natural that the averages go back to the value of f at x. One cannot literally convolve a general $f \in L^p$ with a Dirac measure (point values of f do not make sense), which is why we study the convergence of $f * \Phi_\varepsilon$ in suitable senses.

This is how we approximate: lots of results in analysis are first proved for regular functions (continuous, compactly supported), and then extended to irregular functions (say L^1) by approximation. If Φ is regular (for instance $\Phi \in C_c^\infty$), then the functions $f * \Phi_\varepsilon$ are very regular by Proposition 19.2 and Remark 19.2, and they approximate the irregular function f. A family $\{\Phi_\varepsilon\}$ used this way is called an *approximation of the identity*.

We need a continuous version of the triangle inequality.

Lemma 19.6 (Minkowski's integral inequality)

Let $1 \leq p \leq \infty$ and let $H(x,w)$ be a measurable function on $\mathbb{R}^d \times \mathbb{R}^d$. Then

$$\left\| \int_{\mathbb{R}^d} H(\,\cdot\,,w)\,\mathrm{d}w \right\|_{L^p} \leq \int_{\mathbb{R}^d} \|H(\,\cdot\,,w)\|_{L^p}\,\mathrm{d}w. \tag{19.46}$$

In words: the L^p norm of an integral is at most the integral of the L^p norms. For finite sums this is the usual triangle inequality (applied to $\sum_j H_j$); the integral version is obtained by replacing the sum over j by an integral in w.

Proof. Exercise 15.2. ■

Theorem 19.7 (Approximations of the identity: L^p-convergence)

Let $\Phi \in L^1\left(\mathbb{R}^d\right)$ with $\int_{\mathbb{R}^d} \Phi(x)\,\mathrm{d}x = A$. Then the following statements hold:

(i) if $f \in L^p\left(\mathbb{R}^d\right)$ with $1 \leq p < \infty$, then

$$\lim_{\varepsilon \to 0} \|f * \Phi_\varepsilon - Af\|_{L^p} = 0; \tag{19.47}$$

(ii) if f is bounded and uniformly continuous, then

$$\lim_{\varepsilon \to 0} \|f * \Phi_\varepsilon - Af\|_{L^\infty} = 0; \tag{19.48}$$

(iii) if $f \in L^\infty\left(\mathbb{R}^d\right)$ and f is continuous in an open set U, then $f * \Phi_\varepsilon \to Af$ uniformly on compact subsets of U as $\varepsilon \to 0$.

Statement (i) says that $f * \Phi_\varepsilon$ approximates Af in L^p when f is merely in L^p (possibly a wild function), and (ii), (iii) say that when f is continuous the approximation is uniform (on the appropriate sets).

Proof. Since $\int \Phi_\varepsilon = A$, we can write

$$Af(x) = \int_{\mathbb{R}^d} f(x)\,\Phi_\varepsilon(y)\,\mathrm{d}y \tag{19.49}$$

(the number $f(x)$ comes out of the integral in y). Hence, with the change of variables $y = \varepsilon w$,

$$f * \Phi_\varepsilon(x) - Af(x) = \int_{\mathbb{R}^d} [f(x-y) - f(x)]\,\Phi_\varepsilon(y)\,\mathrm{d}y \tag{19.50}$$

$$= \int_{\mathbb{R}^d} [f(x-\varepsilon w) - f(x)]\,\Phi(w)\,\mathrm{d}w \tag{19.51}$$

$$= \int_{\mathbb{R}^d} [\tau_{\varepsilon w} f(x) - f(x)]\,\Phi(w)\,\mathrm{d}w. \tag{19.52}$$

(i) We take the L^p norm in x of (19.52) and use Minkowski's integral inequality, with

$$H(x,w) = [\tau_{\varepsilon w} f(x) - f(x)]\,\Phi(w) \tag{19.53}$$

to get

$$\|f * \Phi_\varepsilon - Af\|_{L^p} \leq \int_{\mathbb{R}^d} \|\tau_{\varepsilon w} f - f\|_{L^p}\,|\Phi(w)|\,\mathrm{d}w. \tag{19.54}$$

We want to send $\varepsilon \to 0$ inside this integral, and we use Dominated Convergence Theorem:

- for each fixed w, $\|\tau_{\varepsilon w} f - f\|_{L^p} \to 0$ as $\varepsilon \to 0$, by the Continuity of translations in L^p (here $p < \infty$ is used);

- the integrand is dominated: since L^p norm is translation invariant,

$$\|\tau_{\varepsilon w} f - f\|_{L^p} \leq 2\|f\|_{L^p}, \tag{19.55}$$

so the integrand is bounded by $2\|f\|_{L^p}|\Phi(w)|$, which is integrable and independent of ε.

Therefore the RHS of (19.54) tends to 0, which proves (i).

(ii) Let $\eta > 0$. Since $\Phi \in L^1$, there is a compact set E (for instance a large closed ball $\overline{B(0,R)}$) such that

$$\int_{E^c} |\Phi| < \eta. \tag{19.56}$$

Since f is uniformly continuous, there is $\varepsilon_0 > 0$ such that if $\varepsilon < \varepsilon_0$, then

$$|\tau_{\varepsilon w} f(x) - f(x)| < \eta \tag{19.57}$$

for all $w \in E$ and all $x \in \mathbb{R}^d$ (indeed, $|\varepsilon w| \leq \varepsilon R$, so it suffices that εR is smaller than the uniform continuity modulus corresponding to η). Then, by (19.52), for $\varepsilon < \varepsilon_0$ and every $x \in \mathbb{R}^d$,

$$\begin{aligned} |f * \Phi_\varepsilon(x) - Af(x)| &\leq \int_{\mathbb{R}^d} |\tau_{\varepsilon w} f(x) - f(x)|\,|\Phi(w)|\,\mathrm{d}w \\ &= \int_E |\tau_{\varepsilon w} f(x) - f(x)|\,|\Phi(w)|\,\mathrm{d}w + \int_{E^c} |\tau_{\varepsilon w} f(x) - f(x)|\,|\Phi(w)|\,\mathrm{d}w \\ &\leq \eta\|\Phi\|_{L^1} + 2\eta\|f\|_{L^\infty}, \end{aligned} \tag{19.58}$$

where in the first integral we used the choice of ε_0, and in the second one $|\tau_{\varepsilon w} f - f| \leq 2\|f\|_{L^\infty}$ together with $\int_{E^c}|\Phi| < \eta$. The bound does not depend on x, so $\|f * \Phi_\varepsilon - Af\|_{L^\infty} \leq \eta(\|\Phi\|_{L^1} + 2\|f\|_{L^\infty})$ for $\varepsilon < \varepsilon_0$. Since $\eta > 0$ is arbitrary, (ii) follows. (Notice that this is the same splitting into a big compact set, where we use uniform continuity, and its complement, where we use that the tail of Φ is small, that we used for the decay in Proposition 19.4.)

(iii) The proof is similar to the one of (ii). Given $\eta > 0$, keep the choice of the compact set $E = \overline{B(0,R)}$ with $\int_{E^c}|\Phi| < \eta$. Let K be a compact subset of U. Since U is open and K is compact, there is $\rho > 0$ such that the closed neighbourhood

$$K_1 = \{x : \operatorname{dist}(x, K) \leq \rho\} \tag{19.59}$$

is a compact subset of U. For $\varepsilon R < \rho$, we have $x - \varepsilon w \in K_1$ whenever $x \in K$ and $w \in E$. As f is continuous in U, it is uniformly continuous in the compact set K_1, and hence, for ε smaller than some ε_0,

$$\sup_{x \in K,\, w \in E} |f(x - \varepsilon w) - f(x)| < \eta. \tag{19.60}$$

Now the estimate (19.58) holds for $x \in K$ (in the second integral we use $|f(x - \varepsilon w) - f(x)| \leq 2\|f\|_{L^\infty}$, valid for almost every w, and for $x \in U$ by continuity), uniformly in $x \in K$. This proves (iii). ■

Remark 19.7. Theorem 19.7 gives convergence in norm (or uniform convergence if f is continuous). The next result, which will be discussed in the next lecture, gives *pointwise* (almost everywhere) convergence: if $\Phi \in L^1\left(\mathbb{R}^d\right)$ with $\int \Phi = A$ and $|\Phi|$ admits a radial non-increasing majorant $\psi \in L^1\left(\mathbb{R}^d\right)$, then, for $f \in L^p\left(\mathbb{R}^d\right)$ with $1 \leq p \leq \infty$,

$$\lim_{\varepsilon \to 0} f * \Phi_\varepsilon(x) = Af(x) \tag{19.61}$$

for all x in the Lebesgue set of f (in particular, for almost every $x \in \mathbb{R}^d$).

§20. 6 November 2026

§20.i. Pointwise convergence of approximations of the identity

Recall that, given $\Phi \in L^1\left(\mathbb{R}^d\right)$ and $\varepsilon > 0$, we write

$$\Phi_\varepsilon(x) = \frac{1}{\varepsilon^d}\Phi\left(\frac{x}{\varepsilon}\right). \tag{20.1}$$

We have seen in Theorem 19.7 that, if $\int_{\mathbb{R}^d}\Phi = A$, then $f * \Phi_\varepsilon \to Af$ in L^p (for $f \in L^p$, $1 \le p < \infty$) and uniformly for suitable continuous f. Today we study the *pointwise* convergence of $f * \Phi_\varepsilon(x)$ to $Af(x)$.

Example 20.1 1. Let

$$\Phi = \frac{1}{m\left(B\left(0,1\right)\right)}\,\chi_{B(0,1)}, \tag{20.2}$$

the normalized characteristic function of the unit ball, so that $\int\Phi = 1$. Then

$$\Phi_\varepsilon = \frac{1}{m\left(B\left(0,\varepsilon\right)\right)}\chi_{B(0,\varepsilon)}, \tag{20.3}$$

and

$$f * \Phi_\varepsilon(x) = \frac{1}{m\left(B\left(0,\varepsilon\right)\right)}\int_{B(0,\varepsilon)} f(x-y)\,\mathrm{d}y \tag{20.4}$$

$$= \frac{1}{m\left(B\left(x,\varepsilon\right)\right)}\int_{B(x,\varepsilon)} f(z)\,\mathrm{d}z \tag{20.5}$$

is exactly the average of f over the ball $B(x,\varepsilon)$. By Lebesgue differentiation theorem, these averages converge to $f(x)$ for almost every x as $\varepsilon \to 0$ (here the radius is called ε instead of r). So for this particular kernel we already proved a.e. convergence.

2. Let $\Phi(x) = e^{-\pi|x|^2}$ be the Gaussian. It is non-negative, integrable, and radially decreasing.

In general, we want to prove a.e. convergence for a much larger class of kernels. A kernel Φ need not be non-negative, and it need not integrate to 1: it may have positive and negative parts, and $A = \int\Phi$ can be any real number (even 0). We do need some control, which is the following.

Definition 20.1 (Radial non-increasing function)**.** A function $\psi : \mathbb{R}^d \to [0,\infty)$ is **radial non-increasing** if

$$\psi(x) = \psi_0(|x|) \tag{20.6}$$

for some non-increasing function $\psi_0 : [0,\infty) \to [0,\infty)$; equivalently,

$$\psi(x) \le \psi(y) \text{ whenever } |x| \ge |y|. \tag{20.7}$$

Note that $\psi(x) = \psi(y)$ whenever $|x| = |y|$ actually follows from (20.7). Indeed, $|x| \leq |y|$ implies $\psi(x) \geq \psi(y)$, and $|x| \geq |y|$ implies $\psi(x) \leq \psi(y)$.

Theorem 20.1 (Approximations of the identity: pointwise convergence)

Let $\Phi \in L^1(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} \Phi(x)\,dx = A$. Assume that there exists a function $\psi \in L^1(\mathbb{R}^d)$ which is non-negative, radial non-increasing, and which majorizes Φ, i.e.

$$|\Phi(x)| \leq \psi(x) \quad \text{for almost every } x. \tag{20.8}$$

Then, for $f \in L^p(\mathbb{R}^d)$ with $1 \leq p \leq \infty$ (including $p = \infty$), we have

$$\lim_{\varepsilon \to 0} f * \Phi_\varepsilon(x) = Af(x) \quad \text{for } x \text{ in the Lebesgue set of } f. \tag{20.9}$$

In particular, this converges pointwise for a.e. $x \in \mathbb{R}^d$.

Remark 20.1 (About the hypothesis)**.** The kernels that one usually meets in practice are themselves radial non-increasing (for instance the Gaussian, or the normalized characteristic function of a ball), and then one can take $\psi = |\Phi|$. But Φ can be wild as long as it has a radial non-increasing integrable majorant. For example, any bounded Φ with compact support, say in $B(0, R)$, has the majorant

$$\psi = \|\Phi\|_{L^\infty} \chi_{B(0,R)} \tag{20.10}$$

and any Φ with $|\Phi(x)| \leq C(1 + |x|)^{-d-1}$ has the majorant $C(1 + |x|)^{-d-1}$.

This is a deep and nice theorem; it is already beyond the usual course in real analysis, and goes in the direction of harmonic analysis. However, we already have all the tools to prove it. Recall the strategy for proving a pointwise convergence result (the one we followed for the Lebesgue differentiation theorem):

1. the result is true for a dense class of functions;
2. there is a boundedness estimate for the associated maximal operator.

Here, (1) holds: if f is continuous with compact support, then f is bounded and uniformly continuous, and Theorem 19.7 (ii) gives

$$f * \Phi_\varepsilon \to Af \text{ uniformly;} \tag{20.11}$$

in particular pointwise. For (2) we need to introduce the associated maximal operator.

Definition 20.2 (Maximal operator associated to a kernel)**.** Let $\Phi \in L^1(\mathbb{R}^d)$. For a function f we define

$$M_\Phi f(x) := \sup_{\varepsilon > 0} (|f| * |\Phi_\varepsilon|)(x). \tag{20.12}$$

We replaced f by $|f|$ and Φ by $|\Phi|$, because the convolution $f * \Phi_\varepsilon$ may have signs and cancellations, and we want to control it by something positive; and instead of letting $\varepsilon \to 0$ we take the

supremum over all $\varepsilon > 0$: this is the worst possible case. Note that $||\Phi|_\varepsilon| = |\Phi_\varepsilon|$, so using $|\Phi_\varepsilon|$ is the same as taking $|\Phi|_\varepsilon$, and that

$$|f * \Phi_\varepsilon(x)| \leq (|f| * |\Phi_\varepsilon|)(x) \leq M_\Phi f(x) \quad \text{for all } \varepsilon > 0. \tag{20.13}$$

Remark 20.2. If $\Phi = \frac{1}{m(B(0,1))}\chi_{B(0,1)}$, then

$$M_\Phi f(x) = \sup_{\varepsilon>0} \frac{1}{m(B(x,\varepsilon))} \int_{B(x,\varepsilon)} |f(z)|\,\mathrm{d}z = Mf(x) \tag{20.14}$$

is exactly the (centered) Hardy–Littlewood maximal function that we studied. So we are doing nothing essentially different. The next result says that the Hardy–Littlewood maximal operator controls all the maximal operators of convolution type associated to radial non-increasing kernels ("one maximal function to rule them all").

§20.ii. Reduction to the Hardy–Littlewood maximal function

Now, we record some observations that would lead to the proof of Theorem 20.1.

Observation 1. Since $|\Phi| \leq \psi$, we also have $|\Phi_\varepsilon| \leq \psi_\varepsilon$ for every $\varepsilon > 0$, and hence

$$M_\Phi f(x) \leq \sup_{\varepsilon>0} (|f| * \psi_\varepsilon)(x) = M_\psi f(x). \tag{20.15}$$

So it is enough to study the maximal operator associated to a kernel ψ which is non-negative and radial non-increasing, like the Hardy–Littlewood maximal function.

Observation 2 (Super-level sets)**.** Let $u \geq 0$ be radial non-increasing and let $t > 0$. The super-level set

$$S_t = \left\{y \in \mathbb{R}^d : u(y) > t\right\} \tag{20.16}$$

is a ball centered at the origin (open or closed, possibly empty). Indeed, u is radial, so S_t is a union of spheres $\{|y| = \rho\}$; and if $y \in S_t$ and $|y'| \leq |y|$, then

$$u(y') \geq u(y) > t, \tag{20.17}$$

so $y' \in S_t$. Therefore $S_t = \{|y| < r_t\}$ or $\{|y| \leq r_t\}$, with $r_t = \sup\{|y| : y \in S_t\}$. If $u \in L^1$, then

$$m(S_t) \leq \frac{1}{t}\|u\|_{L^1} < \infty, \tag{20.18}$$

by Chebyshev's inequality, so $r_t < \infty$; and since spheres have measure zero, $m(S_t) = m(B(0, r_t))$.

Lemma 20.2

Let $\psi \in L^1(\mathbb{R}^d)$ be non-negative and radial non-increasing. Then for every $h \geq 0$ measurable,

every $x \in \mathbb{R}^d$ and every $\varepsilon > 0$,

$$(h * \psi_\varepsilon)(x) \leq \|\psi\|_{L^1} \, Mh(x), \tag{20.19}$$

where M is the Hardy–Littlewood maximal function. In particular, $M_\psi f \leq \|\psi\|_{L^1} Mf$.

Proof. Let $u = \psi_\varepsilon$. This function is again non-negative, radial and non-increasing, and $\|u\|_{L^1} = \|\psi\|_{L^1}$ (change of variables).

The idea is to make the super-level sets appear using The layer-cake formula. Every non-negative number can be written as an integral of the constant function 1 over an interval, so, for each y,

$$u(y) = \int_0^{u(y)} 1 \, \mathrm{d}t = \int_0^\infty \chi_{\{t<u(y)\}} \, \mathrm{d}t = \int_0^\infty \chi_{S_t}(y) \, \mathrm{d}t. \tag{20.20}$$

Substituting into the convolution and using Tonelli's theorem to switch the integrals (everything is non-negative),

$$(h * u)(x) = \int_{\mathbb{R}^d} h(x-y) \int_0^\infty \chi_{S_t}(y) \, \mathrm{d}t\mathrm{d}y = \int_0^\infty \left(\int_{S_t} h(x-y) \, \mathrm{d}y \right) \mathrm{d}t. \tag{20.21}$$

For fixed t such that $m(S_t) > 0$ the set S_t is a ball centered at the origin, which is symmetric, so with $z = x - y$ the inner integral is an integral of h over a ball centered at x (up to a null set). We multiply and divide by the measure of this ball, to see an *average*:

$$\int_{S_t} h(x-y) \, \mathrm{d}y = m(S_t) \, \frac{1}{m(B(x, r_t))} \int_{B(x,r_t)} h(z) \, \mathrm{d}z \tag{20.22}$$

$$\leq m(S_t) \, Mh(x), \tag{20.23}$$

because the average over a ball centered at x is bounded by the Hardy–Littlewood maximal function. (If $m(S_t) = 0$, the inner integral is 0.) The factor $Mh(x)$ does not depend on t, so

$$(h * u)(x) \leq Mh(x) \int_0^\infty m(S_t) \, \mathrm{d}t = Mh(x) \, \|u\|_{L^1} = \|\psi\|_{L^1} Mh(x). \tag{20.24}$$

Here we used The layer-cake formula to deduce $\|u\|_{L^1} = \int_0^\infty m(\{u > t\}) \, \mathrm{d}t$. ∎

Proposition 20.3 (Weak-type $(1,1)$ estimate for M_Φ)

Under the assumptions of Theorem 20.1 there is a constant C_d, depending only on the dimension, such that for every $f \in L^1(\mathbb{R}^d)$ and every $\lambda > 0$,

$$m(\{x : M_\Phi f(x) > \lambda\}) \leq \frac{C_d \, \|\psi\|_{L^1}}{\lambda} \|f\|_{L^1}. \tag{20.25}$$

Proof. If $\|\psi\|_{L^1} = 0$, then $\Phi = 0$ almost everywhere and there is nothing to prove. Otherwise,

by Observation 1 and Lemma 20.2, $M_\Phi f \le M_\psi f \le \|\psi\|_{L^1} Mf$, so

$$\{M_\Phi f > \lambda\} \subseteq \left\{ Mf > \frac{\lambda}{\|\psi\|_{L^1}} \right\}, \tag{20.26}$$

and by the Weak-type $(1,1)$ estimate for the Hardy–Littlewood maximal function (Theorem 16.6, with constant C_d), the measure of the right-hand set is at most $\frac{C_d \|\psi\|_{L^1}}{\lambda} \|f\|_{L^1}$. So

$$m\left(\{M_\Phi f > \lambda\}\right) \le m\left(\left\{ Mf > \frac{\lambda}{\|\psi\|_{L^1}} \right\}\right) \le \frac{C_d \|\psi\|_{L^1}}{\lambda} \|f\|_{L^1}. \tag{20.27}$$

■

§20.iii. Proof of Approximations of the identity: pointwise convergence

We have now the two ingredients, and all we have to do is to mimic the proof of the Lebesgue differentiation theorem.

Proof of Theorem 20.1. **Step 1: the case $f \in L^1\left(\mathbb{R}^d\right)$.** Define

$$\Omega f(x) := \limsup_{\varepsilon \to 0} |f * \Phi_\varepsilon(x) - Af(x)|. \tag{20.28}$$

We want to show that $\Omega f = 0$ almost everywhere. Fix $\lambda > 0$ and $\eta > 0$. Since $C_c\left(\mathbb{R}^d\right)$ is dense in L^1, choose $g \in C_c\left(\mathbb{R}^d\right)$ with

$$\|f - g\|_{L^1} < \eta. \tag{20.29}$$

By linearity,

$$f * \Phi_\varepsilon - Af = (f-g) * \Phi_\varepsilon - A(f-g) + (g * \Phi_\varepsilon - Ag). \tag{20.30}$$

The last term tends to 0 (even uniformly) as $\varepsilon \to 0$, because g is bounded and uniformly continuous (Theorem 19.7 (ii)): this is the result for the dense class. For the first two terms we use (20.13), which gives, uniformly in ε,

$$|(f-g) * \Phi_\varepsilon(x)| \le M_\Phi(f-g)(x). \tag{20.31}$$

Hence, for every x,

$$\Omega f(x) \le M_\Phi(f-g)(x) + |A|\,|f(x) - g(x)|. \tag{20.32}$$

If $\Omega f(x) > 2\lambda$, then one of the two terms on the right-hand side is larger than λ. Thus

$$\{\Omega f > 2\lambda\} \subseteq \{M_\Phi(f-g) > \lambda\} \cup \{|A|\,|f-g| > \lambda\}. \tag{20.33}$$

By Weak-type $(1,1)$ estimate for M_Φ (applied to $f - g \in L^1$),

$$m^*\left(\{M_\Phi(f-g) > \lambda\}\right) \le \frac{C_d \|\psi\|_{L^1}}{\lambda} \|f-g\|_{L^1} \le \frac{C_d \|\psi\|_{L^1}}{\lambda}\, \eta. \tag{20.34}$$

(We have used outer measure because we don't know if $M_\Phi(f-g)$ is measurable.) By Chebyshev's inequality,

$$m\left(\{|A|\,|f-g|>\lambda\}\right) \leq \frac{|A|}{\lambda}\,\|f-g\|_{L^1} \leq \frac{|A|}{\lambda}\,\eta. \tag{20.35}$$

Therefore, with (20.33), (20.34), (20.35), we have

$$m^*\left(\{\Omega f>2\lambda\}\right) \leq \frac{C_d\,\|\psi\|_{L^1}+|A|}{\lambda}\,\eta. \tag{20.36}$$

The RHS does not depend on g or η, and $\eta>0$ is arbitrary, so $m\left(\{\Omega f>2\lambda\}\right)=0$ for every $\lambda>0$. Taking $\lambda=\frac{1}{n}$, $n\in\mathbb{Z}_{>0}$, and a countable union, $\Omega f=0$ almost everywhere, i.e. $f*\Phi_\varepsilon(x)\to Af(x)$ for almost every x.

Step 2: the general case $f\in L^p$, $1\leq p\leq\infty$. The statement is local, so it suffices to prove convergence for almost every x in $B(0,N)$, for each fixed $N\in\mathbb{Z}_{>0}$. Write $f=f_1+f_2$, with

$$f_1=f\,\chi_{B(0,2N)}, \qquad f_2=f\,\chi_{\mathbb{R}^d\setminus B(0,2N)}. \tag{20.37}$$

By Hölder's inequality,

$$\int_{\mathbb{R}^d}|f_1| \leq \|f\|_{L^p}\,m\left(B(0,2N)\right)^{1-\frac{1}{p}}<\infty, \tag{20.38}$$

so $f_1\in L^1$ and Step 1 applies to f_1:

$$f_1*\Phi_\varepsilon(x)\to Af_1(x) \qquad \text{for a.e. } x; \tag{20.39}$$

and $f_1(x)=f(x)$ for $x\in B(0,N)$. It remains to show that $f_2*\Phi_\varepsilon\to 0$ (uniformly) on $B(0,N)$. Let $x\in B(0,N)$. If $f_2(x-y)\neq 0$, then $|x-y|\geq 2N$, so

$$|y|\geq|x-y|-|x|>N. \tag{20.40}$$

Using $|\Phi_\varepsilon|\leq\psi_\varepsilon$ and Hölder's inequality (with $\frac{1}{p}+\frac{1}{p'}=1$),

$$|f_2*\Phi_\varepsilon(x)| \leq \int_{|y|>N}|f(x-y)|\,\psi_\varepsilon(y)\,\mathrm{d}y \tag{20.41}$$

$$\leq \|f\|_{L^p}\left\|\psi_\varepsilon\chi_{\{|y|>N\}}\right\|_{L^{p'}}. \tag{20.42}$$

We show that the right-hand side tends to 0 as $\varepsilon\to 0$, by the following lemma (Lemma 20.4) on the tails of radial non-increasing functions. This finishes the proof, since then, for a.e. $x\in B(0,N)$,

$$f*\Phi_\varepsilon(x)-Af(x)=\left[f_1*\Phi_\varepsilon(x)-Af_1(x)\right]+f_2*\Phi_\varepsilon(x)\to 0. \tag{20.43}$$

Lemma 20.4 (Tails of radial non-increasing kernels)

Let $\psi \in L^1\left(\mathbb{R}^d\right)$ be non-negative and radial non-increasing, let $N > 0$, and for $s > 0$ put $\theta(s) = \int_{|z| \geq s} \psi(z)\, \mathrm{d}z$, so that $\theta(s) \to 0$ as $s \to \infty$. Then, as $\varepsilon \to 0$:

(a)
$$\int_{|y|>N} \psi_\varepsilon(y)\, \mathrm{d}y \leq \theta\left(\frac{N}{\varepsilon}\right) \to 0; \tag{20.44}$$

(b)
$$\sup_{|y|>N} \psi_\varepsilon(y) \leq \frac{1}{c_d N^d}\, \theta\left(\frac{N}{2\varepsilon}\right) \to 0, \tag{20.45}$$

where $c_d = \left(1 - 2^{-d}\right) m\left(B(0,1)\right)$;

(c) for every $1 \leq q \leq \infty$, $\left\|\psi_\varepsilon \chi_{\{|y|>N\}}\right\|_{L^q} \to 0$.

Proof. (a) The change of variables $y = \varepsilon z$ gives

$$\int_{|y|>N} \psi_\varepsilon(y)\, \mathrm{d}y = \int_{|z|>\frac{N}{\varepsilon}} \psi(z)\, \mathrm{d}z \leq \theta\left(\frac{N}{\varepsilon}\right), \tag{20.46}$$

which tends to 0 since $\psi \in L^1$ (Dominated Convergence Theorem).

(b) Let $|y| > N$ and $z = \frac{y}{\varepsilon}$, $R = |z| > \frac{N}{\varepsilon}$. For every w in the annulus $\left\{\frac{R}{2} \leq |w| \leq R\right\}$ we have $\psi(w) \geq \psi(z)$, since ψ is radial non-increasing. The annulus has measure $c_d R^d$, so

$$\psi(z)\, c_d R^d \leq \int_{\frac{R}{2} \leq |w| \leq R} \psi(w)\, \mathrm{d}w \leq \theta\left(\frac{R}{2}\right) \leq \theta\left(\frac{N}{2\varepsilon}\right), \tag{20.47}$$

because θ is non-increasing. Since $R > \frac{N}{\varepsilon}$, we get

$$\psi_\varepsilon(y) = \varepsilon^{-d} \psi(z) \leq \varepsilon^{-d}\, \frac{\theta\left(\frac{N}{2\varepsilon}\right)}{c_d \left(\frac{N}{\varepsilon}\right)^d} = \frac{1}{c_d N^d} \theta\left(\frac{N}{2\varepsilon}\right). \tag{20.48}$$

(c) For $q = \infty$ this is (b), and for $q = 1$ this is (a). For $1 < q < \infty$,

$$\int_{|y|>N} \psi_\varepsilon^q \leq \left(\sup_{|y|>N} \psi_\varepsilon\right)^{q-1} \int_{|y|>N} \psi_\varepsilon \to 0 \tag{20.49}$$

by (a) and (b). ∎

By Lemma 20.4 (c) with $q = p'$, the right-hand side of (20.42) tends to 0 as $\varepsilon \to 0$, uniformly in $x \in B(0, N)$, which completes the proof of Theorem 20.1. ∎

In general, we cannot do better than almost everywhere. If f is not continuous, the convergence can fail at some points.

Remark 20.3 (Other families of sets)**.** The theorem applies to $\Phi = \frac{1}{m(E)} \chi_E$ for any bounded measurable set E with $m(E) > 0$ (for instance a cube, or a set containing the origin which is

not convex). Because, if $E \subseteq B(0,R)$, then

$$\Phi \leq \frac{1}{m(E)} \chi_{B(0,R)}, \tag{20.50}$$

which is a radial non-increasing integrable majorant. Then $f * \Phi_\varepsilon(x)$ is the average of f over the dilated and reflected set $x - \varepsilon E$, and Theorem 20.1 says that these averages converge to $f(x)$ for almost every x, in the same way as averages over balls.

Remark 20.4 (Why not Dominated Convergence Theorem?)**.** For fixed x one could try to pass to the limit in $\int f(x-y)\,\Phi_\varepsilon(y)\,dy$ using the Dominated Convergence Theorem. But, for instance for $\Phi = \frac{1}{2}\chi_{[-1,1]}$ in dimension 1, the height of Φ_ε is $\frac{1}{2\varepsilon}$, which blows up as $\varepsilon \to 0$: the family is not uniformly bounded, so there is no dominating function, and indeed the pointwise limit of $\Phi_\varepsilon(y)$ is 0 for every $y \neq 0$, whereas the integral $\int \Phi_\varepsilon$ stays equal to 1. The mass concentrates at the single point $y = 0$ (this is the convergence to a Dirac delta, a convergence of *measures*). That is why a pointwise convergence result of this type is more sophisticated than an application of Dominated Convergence Theorem, and why maximal functions are needed.

§20.iv. The Fourier transform

Until now, in many arguments we controlled an object by its modulus: we moved the absolute value inside the integral and controlled everything by positive quantities. This is not always possible. In many problems in analysis one really has to understand the *oscillation* of the objects. A basic example is the integral

$$\int_0^\infty \frac{\sin x}{x} dx. \tag{20.51}$$

The function $\frac{\sin x}{x}$ is not Lebesgue integrable on $(0,\infty)$, so we cannot just move the modulus inside (the integral of the modulus is infinite). But the limit $\lim_{T\to\infty} \int_0^T \frac{\sin x}{x} dx$ exists (it equals $\frac{\pi}{2}$), because the integrand has cancellation. To understand such limits one has to understand the cancellation. The easiest functions with cancellation are the sines and cosines: $\sin x, \sin 2x, \sin 3x, \ldots$

This idea goes back to Fourier (early 1800s), who wrote a periodic function (say on the interval $[0,1]$) as a sum of sines and cosines, the basic functions: if one understands each of the parts separately, one hopes to understand the function as a whole. Historically, the theory of Fourier series came before the theory of the Fourier transform, which is the same theory but on $\mathbb{R}^d$ instead of a periodic setting. We start with $\mathbb{R}^d$.

Definition 20.3 (Fourier transform)**.** Let $f \in L^1(\mathbb{R}^d)$. The **Fourier transform** of f is the function $\widehat{f} : \mathbb{R}^d \to \mathbb{C}$ defined by

$$\widehat{f}(\xi) := \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i x\cdot\xi}\, dx, \tag{20.52}$$

where $x \cdot \xi$ is the inner product in $\mathbb{R}^d$.

So $\widehat{f}(\xi)$ is the integral of f against the oscillatory kernel $e^{-2\pi i x\cdot\xi}$, which oscillates very fast when ξ is large. Since $\left|e^{-2\pi i x\cdot\xi}\right| = 1$, the integral is well defined for every ξ when $f \in L^1$. (Different books use a plus sign, or no 2π, in the exponent; this is just a matter of normalization, and we pick the normalization with $e^{-2\pi i x\cdot\xi}$.)

Remark 20.5 (Fourier transform, derivatives and ODEs)**.** Why is this object useful? An initial motivation: it interacts very well with derivatives. Suppose, for simplicity, that f is smooth with compact support and look at the difference quotient of $\widehat{f}$ in the variable ξ_j:

$$\frac{\widehat{f}(\xi + te_j) - \widehat{f}(\xi)}{t} = \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i x\cdot\xi}\, \frac{e^{-2\pi i t x_j} - 1}{t} \mathrm{d}x. \tag{20.53}$$

To let $t \to 0$ inside the integral we use Dominated Convergence Theorem: the mean value theorem bounds the quotient by $2\pi |x_j|$, which is bounded on the (compact) support of f. We get

$$\partial_{\xi_j} \widehat{f}(\xi) = \int_{\mathbb{R}^d} (-2\pi i x_j)\, f(x)\, e^{-2\pi i x\cdot\xi} \mathrm{d}x = \widehat{(-2\pi i x_j f)}(\xi) : \tag{20.54}$$

a derivative on one side becomes a multiplication by x_j on the other side. Dually, multiplying $\widehat{f}$ by $2\pi i\xi_j$ corresponds to differentiating f (by integration by parts, see Proposition 20.6 (vi) below).

This is particularly useful to solve differential equations. For instance, to solve the ODE

$$f''' + 5f'' + 3f' + f = g \tag{20.55}$$

(the coefficients are only illustrative), we take the Fourier transform on both sides. The derivatives become multiplications by $2\pi i\xi$:

$$\left((2\pi i\xi)^3 + 5(2\pi i\xi)^2 + 3(2\pi i\xi) + 1\right) \widehat{f}(\xi) = \widehat{g}(\xi), \tag{20.56}$$

and therefore, wherever the polynomial does not vanish, $\widehat{f}$ is equal to $\widehat{g}$ divided by this polynomial. If we find the Fourier transform of the function, we find the function by coming back. Thus the Fourier transform turns derivatives into multiplications by polynomials.

§20.v. Basic properties of Fourier transform

We gather several properties in one proposition.

Notation.

- For $y \in \mathbb{R}^d$, $\tau_y f(x) = f(x - y)$ is the translation.
- A **multi-index** is $\alpha = (\alpha_1, \ldots, \alpha_d)$ with $\alpha_i \in \mathbb{Z}_{\geq 0}$, its **order** is

$$|\alpha| = \alpha_1 + \cdots + \alpha_d. \tag{20.57}$$

- For $x \in \mathbb{R}^d$ we write

$$x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d} \tag{20.58}$$

(a polynomial), and

$$\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d} \tag{20.59}$$

(a derivative operator).

Lemma 20.5

If $h \in L^1\left(\mathbb{R}^d\right)$, then $\widehat{h}$ is a continuous function.

Proof. Let $\xi_n \to \xi$. Then $h(x)\, e^{-2\pi i x\cdot\xi_n} \to h(x)\, e^{-2\pi i x\cdot\xi}$ for every x, and the functions are dominated by $|h| \in L^1$. By Dominated Convergence Theorem, $\widehat{h}(\xi_n) \to \widehat{h}(\xi)$. ∎

Proposition 20.6 (Basic properties of the Fourier transform)

Let $f, g \in L^1\left(\mathbb{R}^d\right)$.

(i) $\left\|\widehat{f}\right\|_{L^\infty} \le \|f\|_{L^1}$.

(ii) For $y \in \mathbb{R}^d$,

$$\widehat{\tau_y f}(\xi) = e^{-2\pi i y\cdot\xi}\, \widehat{f}(\xi). \tag{20.60}$$

(iii) Let $T : \mathbb{R}^d \to \mathbb{R}^d$ be an invertible linear transformation. Then

$$\widehat{f\circ T}(\xi) = \frac{1}{|\det T|}\, \widehat{f}\left((T^*)^{-1}\xi\right), \tag{20.61}$$

where T^* is the adjoint (transpose) of T. In particular, if T is a rotation (or any orthogonal transformation), then $\widehat{f\circ T} = \widehat{f}\circ T$; and if $Tx = \lambda x$ is a dilation ($\lambda > 0$), then

$$\widehat{f(\lambda\,\cdot\,-)}(\xi) = \frac{1}{\lambda^d}\widehat{f}\left(\frac{\xi}{\lambda}\right). \tag{20.62}$$

(iv) $\widehat{f * g} = \widehat{f}\,\widehat{g}$.

(v) Let $k \in \mathbb{Z}_{>0}$. If $x^\alpha f \in L^1\left(\mathbb{R}^d\right)$ for all multi-indices α with $|\alpha| \le k$, then $\widehat{f} \in C^k\left(\mathbb{R}^d\right)$ and for $|\alpha| \le k$,

$$\partial^\alpha \widehat{f}(\xi) = \widehat{\left((-2\pi i x)^\alpha f\right)}(\xi). \tag{20.63}$$

(vi) Let $k \in \mathbb{Z}_{>0}$. If $f \in C^k\left(\mathbb{R}^d\right)$, $\partial^\alpha f \in L^1\left(\mathbb{R}^d\right)$ for $|\alpha| \le k$, and $\partial^\alpha f \in C_0\left(\mathbb{R}^d\right)$ for $|\alpha| \le k-1$, then for $|\alpha| \le k$,

$$\widehat{\partial^\alpha f}(\xi) = (2\pi i\xi)^\alpha\, \widehat{f}(\xi). \tag{20.64}$$

In words: translations on one side become modulations (multiplications by exponentials) on the other side; composition with a linear transformation is a change of variables; convolutions become products; multiplication by polynomials becomes differentiation; and differentiation becomes multiplication by polynomials. The notation: here $(-2\pi i x)^\alpha = (-2\pi i)^{|\alpha|}\, x^\alpha$ and $(2\pi i\xi)^\alpha = (2\pi i)^{|\alpha|}\, \xi^\alpha$.

Proof. (i) Moving the absolute value inside the integral and using $\left|e^{-2\pi i x\cdot\xi}\right| = 1$, for every ξ,

$$\left|\widehat{f}(\xi)\right| \leq \int_{\mathbb{R}^d} |f(x)|\,\mathrm{d}x = \|f\|_{L^1}. \tag{20.65}$$

So the supremum norm of the Fourier transform is controlled by the L^1 norm: this is the most basic inequality.

(ii) With the change of variables $z = x - y$ (so $x = z + y$),

$$\widehat{\tau_y f}(\xi) = \int_{\mathbb{R}^d} f(x-y)\, e^{-2\pi i x\cdot\xi}\mathrm{d}x = \int_{\mathbb{R}^d} f(z)\, e^{-2\pi i (z+y)\cdot\xi}\mathrm{d}z = e^{-2\pi i y\cdot\xi}\widehat{f}(\xi). \tag{20.66}$$

By the same computation,

$$\widehat{\left(e^{2\pi i x\cdot\eta} f\right)}(\xi) = \widehat{f}(\xi-\eta). \tag{20.67}$$

(iii) This is again a change of variables. Let $z = Tx$, so $x = T^{-1}z$ and $\mathrm{d}x = \frac{1}{|\det T|}\mathrm{d}z$. Since

$$T^{-1}z\cdot\xi = z\cdot\left(T^{-1}\right)^*\xi = z\cdot\left(T^*\right)^{-1}\xi, \tag{20.68}$$

we get

$$\widehat{f\circ T}(\xi) = \int_{\mathbb{R}^d} f(Tx)\, e^{-2\pi i x\cdot\xi}\mathrm{d}x = \frac{1}{|\det T|}\int_{\mathbb{R}^d} f(z)\, e^{-2\pi i z\cdot\left(T^*\right)^{-1}\xi}\mathrm{d}z = \frac{1}{|\det T|}\widehat{f}\left(\left(T^*\right)^{-1}\xi\right). \tag{20.69}$$

If T is orthogonal, $TT^* = I$, then $|\det T| = 1$ and $\left(T^*\right)^{-1} = T$, which gives $\widehat{f\circ T} = \widehat{f}\circ T$. For $T = \lambda I$ we have $|\det T| = \lambda^d$ and $\left(T^*\right)^{-1}\xi = \frac{\xi}{\lambda}$.

(iv) By Tonelli's theorem,

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(x-y)|\,|g(y)|\,\mathrm{d}y\mathrm{d}x = \|f\|_{L^1}\|g\|_{L^1} < \infty, \tag{20.70}$$

so $f * g \in L^1$ and Fubini's theorem can be used. We split the exponential as

$$e^{-2\pi i x\cdot\xi} = e^{-2\pi i (x-y)\cdot\xi} e^{-2\pi i y\cdot\xi}, \tag{20.71}$$

so that

$$\widehat{f*g}(\xi) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} f(x-y)\, e^{-2\pi i (x-y)\cdot\xi}\, g(y)\, e^{-2\pi i y\cdot\xi}\mathrm{d}y\mathrm{d}x. \tag{20.72}$$

Exchanging the order of integration, the inner integral in x is (after the change of variables $z = x - y$) equal to $\widehat{f}(\xi)$, independently of y, and then

$$\widehat{f*g}(\xi) = \widehat{f}(\xi)\int_{\mathbb{R}^d} g(y)\, e^{-2\pi i y\cdot\xi}\mathrm{d}y = \widehat{f}(\xi)\,\widehat{g}(\xi). \tag{20.73}$$

(v) *Step a: one derivative.* Let $h \in L^1$ with $x_j h \in L^1$. Then

$$\frac{\widehat{h}(\xi + te_j) - \widehat{h}(\xi)}{t} = \int_{\mathbb{R}^d} h(x)\, e^{-2\pi i x\cdot\xi}\,\frac{e^{-2\pi i t x_j} - 1}{t}\mathrm{d}x. \tag{20.74}$$

As $t \to 0$, the quotient $\frac{e^{-2\pi i t x_j}-1}{t}$ tends to $-2\pi i x_j$ for each x, and, since $\left|e^{i\theta}-1\right| \leq |\theta|$ (mean value theorem), it is bounded by $2\pi |x_j|$. Hence the integrand is dominated by $2\pi |x_j| |h(x)|$, an L^1 function independent of t. By the Dominated Convergence Theorem, $\partial_{\xi_j} \widehat{h}$ exists and

$$\partial_{\xi_j} \widehat{h}(\xi) = \int_{\mathbb{R}^d} (-2\pi i x_j) h(x) e^{-2\pi i x \cdot \xi} \mathrm{d}x = \widehat{(-2\pi i x_j h)}(\xi). \tag{20.75}$$

Step b: induction. We prove $\partial^\alpha \widehat{f} = \widehat{((-2\pi i x)^\alpha f)}$ by induction on $|\alpha| \leq k$. The case $\alpha = 0$ is trivial. If it is true for α with $|\alpha| < k$, put $h = (-2\pi i x)^\alpha f \in L^1$. Then $x_j h$ is a multiple of $x^{\alpha+e_j} f$, which is in L^1 because $|\alpha + e_j| \leq k$. Step a gives

$$\partial_{\xi_j} \partial^\alpha \widehat{f} = \widehat{(-2\pi i x_j h)} = \widehat{\left((-2\pi i x)^{\alpha+e_j} f\right)}. \tag{20.76}$$

Finally, all the partial derivatives of $\widehat{f}$ of order $\leq k$ are Fourier transforms of L^1 functions, hence continuous by Lemma 20.5, so $\widehat{f} \in C^k$. (This is the Dominated Convergence Theorem computation of the previous remark.)

(vi) It suffices to treat one derivative, $\alpha = e_j$, and then to iterate: the hypotheses for $\partial^\beta f$, $|\beta| \leq k-1$, say that $\partial^\beta f \in C^1$ with $\partial^\beta f$ and $\partial_j \partial^\beta f$ in L^1 (and $\partial^\beta f \in C_0$), so the one-derivative statement can be applied to $\partial^\beta f$ successively, giving the factor $2\pi i \xi_j$ at each step.

So let $u \in C^1\left(\mathbb{R}^d\right)$ with $u, \partial_j u \in L^1\left(\mathbb{R}^d\right)$ and $u \in C_0$. We use integration by parts in the variable x_j, which is how derivatives are moved from u to the exponential. Write $x = (x', t)$ with $t = x_j$ (and x' the remaining variables), and $\xi = (\xi', \xi_j)$. By Fubini's theorem, for almost every x', the functions $t \mapsto u(x', t)$ and $t \mapsto \partial_j u(x', t)$ belong to $L^1(\mathbb{R})$. For such x', the integration by parts formula on $[-R, R]$ gives

$$\int_{-R}^{R} \partial_j u(x', t) e^{-2\pi i t \xi_j} \mathrm{d}t = \left[u(x', t) e^{-2\pi i t \xi_j}\right]_{t=-R}^{t=R} + 2\pi i \xi_j \int_{-R}^{R} u(x', t) e^{-2\pi i t \xi_j} \mathrm{d}t. \tag{20.77}$$

As $R \to \infty$, the boundary term tends to 0: indeed

$$u(x', t) = u(x', 0) + \int_0^t \partial_j u(x', s) \, \mathrm{d}s \tag{20.78}$$

has limits as $t \to \pm\infty$ (because $\partial_j u(x', \cdot) \in L^1$), and these limits must vanish because $u(x', \cdot) \in L^1(\mathbb{R})$ (this is where the decay of u enters; note that the integrability of u and $\partial_j u$ already guarantees that the boundary terms vanish, so the extra assumption $u \in C_0$ is not needed in this argument). Hence

$$\int_{\mathbb{R}} \partial_j u(x', t) e^{-2\pi i t \xi_j} \mathrm{d}t = 2\pi i \xi_j \int_{\mathbb{R}} u(x', t) e^{-2\pi i t \xi_j} \mathrm{d}t \quad \text{for almost every } x'. \tag{20.79}$$

Multiplying by $e^{-2\pi i x' \cdot \xi'}$ and integrating in x' (Fubini's theorem applies since $u, \partial_j u \in L^1$), we obtain $\widehat{\partial_j u}(\xi) = 2\pi i \xi_j \widehat{u}(\xi)$. ∎

Remark 20.6 (Summary)**.** To summarize the properties of the Fourier transform with respect

to the basic operations:

1. translations in x correspond to modulations (multiplication by $e^{-2\pi i y\cdot\xi}$) in ξ, and vice versa;
2. composition with a linear map is a change of variables on the Fourier side (rotations on one side give rotations on the other, dilations give inverse dilations);
3. convolutions become products;
4. multiplication by polynomials becomes differentiation on the other side, and differentiation becomes multiplication by polynomials.

§20.vi. Exercises

Exercise 20.1. Let $f(x) = \frac{1}{|x|^2(1+|x|^2)}$ for $x \in \mathbb{R}^3$, with the usual Lebesgue measure.

(i) For which values of $p \in (0,\infty]$, do we have $f \in L^p\left(\mathbb{R}^3\right)$?

(ii) Let $g \in L^2\left(\mathbb{R}^3\right)$. Prove that $f * g \in L^2\left(\mathbb{R}^3\right)$ and determine an explicit constant C such that

$$\|f * g\|_2 \leq C \|g\|_2 . \tag{20.80}$$

Solution. (i) First, note that $f \notin L^\infty\left(\mathbb{R}^3\right)$, since

$$\lim_{x\to 0} f(x) = \infty, \tag{20.81}$$

so f isn't bounded on any interval around 0. So f can't be essentially bounded. For $p < \infty$, $f \in L^p$ if and only if $\int_{\mathbb{R}^3} f^p < \infty$. We convert this integral into polar coordinates:

$$\int_{\mathbb{R}^3} f^p = 4\pi \int_0^\infty \frac{r^2}{r^{2p}\,(1+r^2)^p}\,\mathrm{d}r \tag{20.82}$$

$$= 4\pi \int_0^1 \frac{1}{r^{2p-2}\,(1+r^2)^p}\,\mathrm{d}r + 4\pi \int_1^\infty \frac{1}{r^{2p-2}\,(1+r^2)^p}\,\mathrm{d}r. \tag{20.83}$$

This is finite if and only if both the integrals are finite. For $0 < r \leq 1$,

$$1 \leq 1 + r^2 \leq 2 \implies 2^{-p} \leq \left(1+r^2\right)^{-p} \leq 1 \tag{20.84}$$

$$\implies \frac{2^{-p}}{r^{2p-2}} \leq \frac{\left(1+r^2\right)^{-p}}{r^{2p-2}} \leq \frac{1}{r^{2p-2}}. \tag{20.85}$$

So we have $\int_0^1 \frac{1}{r^{2p-2}(1+r^2)^p}\,\mathrm{d}r < \infty$ if and only if $\int_0^1 \frac{1}{r^{2p-2}}\,\mathrm{d}r < \infty$.

$$\int_0^1 \frac{1}{r^{2p-2}}\,\mathrm{d}r = \left[\frac{r^{2-2p+1}}{2-2p+1}\right]_0^1 = \frac{1}{3-2p} - \frac{1}{3-2p}\lim_{r\to 0^+} r^{3-2p}, \tag{20.86}$$

which is finite if and only if $3-2p>0$, i.e. $p<\frac{3}{2}$. Now,

$$\frac{1}{r^{2p-2}\left(1+r^2\right)^p}=\frac{1}{r^{4p-2}\left(1+\frac{1}{r^2}\right)^p}. \tag{20.87}$$

So for $r\geq 1$,

$$1\leq 1+\frac{1}{r^2}\leq 2 \implies 2^{-p}\leq\left(1+\frac{1}{r^2}\right)^{-p}\leq 1 \tag{20.88}$$

$$\implies \frac{2^{-p}}{r^{4p-2}}\leq\frac{\left(1+\frac{1}{r^2}\right)^{-p}}{r^{4p-2}}\leq\frac{1}{r^{4p-2}}. \tag{20.89}$$

So we have $\int_1^\infty \frac{1}{r^{2p-2}(1+r^2)^p}\,\mathrm{d}r<\infty$ if and only if $\int_1^\infty \frac{1}{r^{4p-2}}\,\mathrm{d}r<\infty$.

$$\int_1^\infty\frac{1}{r^{4p-2}}\,\mathrm{d}r=\left[\frac{r^{2-4p+1}}{2-4p+1}\right]_1^\infty=\frac{1}{3-4p}\lim_{r\to\infty}r^{3-4p}-\frac{1}{3-4p}, \tag{20.90}$$

which is finite if and only if $3-4p<0$, i.e. $p>\frac{3}{4}$. Therefore, $f\in L^p$ for $p\in\left(\frac{3}{4},\frac{3}{2}\right)$.

(ii) Young's inequality says that if $1+\frac{1}{r}=\frac{1}{p}+\frac{1}{q}$, then

$$\|f*g\|_r\leq\|f\|_p\|g\|_q. \tag{20.91}$$

Taking $q=r=2$, we must have that $p=1$. Since $f\in L^1$, so we can apply Young's inequality. Then we have

$$\|f*g\|_2\leq\|f\|_1\|g\|_2, \tag{20.92}$$

so the constant $C=\|f\|_1$. Let's evaluate $\|f\|_1$.

$$\|f\|_1=\int_{\mathbb{R}^3}|f|=4\pi\int_0^\infty\frac{r^2}{r^2\left(1+r^2\right)}\,\mathrm{d}r=4\pi\int_0^\infty\frac{1}{\left(1+r^2\right)}\,\mathrm{d}r \tag{20.93}$$

$$=4\pi\lim_{r\to\infty}\arctan(r)=4\pi\frac{\pi}{2}=2\pi^2. \tag{20.94}$$

So $C=2\pi^2$.

■

Exercise 20.2. Let $f\in L^2\left(\mathbb{R}^2\right)$ and define

$$g_n(x)=(\phi_n*f)(x)=\int_{\mathbb{R}^2}\phi_n(x-y)f(y)\,\mathrm{d}y,\qquad \phi_n(x)=n^4x_1x_2e^{-n|x|}, \tag{20.95}$$

for all $x=(x_1,x_2)\in\mathbb{R}^2$, where $|x|=\sqrt{x_1^2+x_2^2}$.

(i) Show that $\{g_n\}$ is a Cauchy sequence in $L^2\left(\mathbb{R}^2\right)$ and determine its limit.

(ii) If $\phi_n(x)=n^4\,|x_1|\,|x_2|\,e^{-n|x|}$ instead, what would be such limit?

Solution. (i) Let $\Phi := \phi_1$. Then first notice that $\Phi \in L^1$.

$$\int_{\mathbb{R}^2} |\Phi| = \int_{\mathbb{R}^2} |xy|\, e^{-\sqrt{x^2+y^2}}\, \mathrm{d}\mathbf{x} \tag{20.96}$$

$$= \int_0^\infty \int_0^{2\pi} r^2\, |\cos\theta \sin\theta|\, e^{-r}\ r\, \mathrm{d}\theta\, \mathrm{d}r \tag{20.97}$$

$$= \left(\int_0^\infty r^3 e^{-r}\, \mathrm{d}r\right)\left(\frac{1}{2}\int_0^{2\pi} |\sin 2\theta|\ \mathrm{d}\theta\right). \tag{20.98}$$

The first integral $\int_0^\infty r^3 e^{-r}\, \mathrm{d}r$ is just $3! = 6$, and the second integral is

$$\int_0^{2\pi} |\sin 2\theta|\ \mathrm{d}\theta = 4\int_0^{\frac{\pi}{2}} \sin 2\theta\, \mathrm{d}\theta = 4\left[-\frac{\cos 2\theta}{2}\right]_0^{\frac{\pi}{2}} = 4. \tag{20.99}$$

Therefore,

$$\int_{\mathbb{R}^2} |\Phi| = 6 \cdot \frac{1}{2} \cdot 4 = 12. \tag{20.100}$$

So $\varphi \in L^1$. Now, by Fubini's theorem,

$$\int_{\mathbb{R}^2} \Phi = \int_{\mathbb{R}^2} xye^{-n\sqrt{x^2+y^2}}\, \mathrm{d}\mathbf{x} = \int_{\mathbb{R}}\int_{\mathbb{R}} xye^{-n\sqrt{x^2+y^2}}\, \mathrm{d}x\mathrm{d}y. \tag{20.101}$$

With a change of variable $x = -u$, we have

$$\int_{\mathbb{R}^2} \Phi = -\int_{\mathbb{R}}\int_{\infty}^{-\infty} -uye^{-n\sqrt{u^2+y^2}}\, \mathrm{d}u\mathrm{d}y \tag{20.102}$$

$$= -\int_{\mathbb{R}}\int_{-\infty}^{\infty} uye^{-n\sqrt{u^2+y^2}}\, \mathrm{d}u\mathrm{d}y = -\int_{\mathbb{R}^2} \Phi. \tag{20.103}$$

So we have

$$\int_{\mathbb{R}^2} \Phi = 0. \tag{20.104}$$

Now, in the same sense as Φ_ε,

$$\Phi_{\frac{1}{n}}(\mathbf{x}) = n^2\Phi(n\mathbf{x}) = n^2\ nx\ ny\ e^{-n|\mathbf{x}|} = \phi_n(\mathbf{x}). \tag{20.105}$$

So we have

$$\int_{\mathbb{R}^2} \phi_n = \int_{\mathbb{R}^2} \Phi = 0 \text{ and } \int_{\mathbb{R}^2} |\phi_n| = \int_{\mathbb{R}^2} |\Phi| = 12. \tag{20.106}$$

Now,

$$g_n = \phi_n * f = \Phi_{\frac{1}{n}} * f, \tag{20.107}$$

and by Approximations of the identity: L^p-convergence, let $A := \int \Phi$,

$$\lim_{n\to\infty} \|g_n - Af\|_2 = \lim_{n\to\infty} \left\|\Phi_{\frac{1}{n}} * f - Af\right\|_2 = 0. \tag{20.108}$$

So g_n is a Cauchy sequence converging to $Af = 0$.

(ii) This is essentially replacing Φ with $|\Phi|$. So $A = \int |\Phi| = 12$. Hence, in this case, g_n is a

Cauchy sequence converging to $Af = 12f$.

■

Exercise 20.3. The Gauss–Weierstrass kernel $K : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$K(x) = \frac{1}{(4\pi)^{\frac{d}{2}}} e^{-\frac{|x|^2}{4}}. \tag{20.109}$$

For $t > 0$ define $K_t(x) = t^{-\frac{d}{2}} K\left(\frac{x}{\sqrt{t}}\right)$. Let $f \in L^p\left(\mathbb{R}^d\right)$, with $1 \leq p \leq \infty$. Prove that $u(x,t) = f * K_t(x) \in C^\infty\left(\mathbb{R}^d \times (0,\infty)\right)$ and verifies

$$u_t - \Delta u = 0 \quad \text{in } \mathbb{R}^d \times (0,\infty), \tag{20.110}$$

with

$$\lim_{t \to 0^+} u(x,t) = f(x) \quad \text{a.e. } x \in \mathbb{R}^d. \tag{20.111}$$

Solution. First we prove a result that allows us to differentiate under the integral sign.

Claim 20.1: Let $F : U \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function for $U \subseteq \mathbb{R}^n$ open, such that

1. $F(x,-)$ is L^1 for every $x \in U$;
2. $F(-,y)$ is C^1 for a.e. $y \in \mathbb{R}^d$;
3. For every compact subset $E \subset U$ and every i, there exists $g_{E,i} \in L^1\left(\mathbb{R}^d\right)$ such that

$$\sup_{x \in E} |\partial_i F(x,y)| \leq g_{E,i}(y) \text{ for a.e. } y. \tag{20.112}$$

Then $I(x) = \int_{\mathbb{R}^d} F(x,y)\, \mathrm{d}y$ is C^1 and $\partial_i I(x) = \int_{\mathbb{R}^d} \partial_i F(x,y)\, \mathrm{d}y$. (Here all the derivatives are with respect to x, and not y.)

Proof. We first show that $\partial_i I$ exists at x. Pick r such that $E := \overline{B(x,r)} \subseteq U$. Then for $|h| < r$,

$$\frac{I(x + he_i) - I(x)}{h} = \int_{\mathbb{R}^d} \frac{F(x + he_i, y) - F(x,y)}{h}\, \mathrm{d}y. \tag{20.113}$$

By mean value theorem,

$$\left|\frac{F(x + he_i, y) - F(x,y)}{h}\right| = |\partial_i F(\xi, y)|, \tag{20.114}$$

for some ξ between x and $x + he_i$. By hypothesis, $|\partial_i F(\xi,y)| \leq g_{E,i}(y)$ a.e. and the dominating function is integrable. So we can apply Dominated Convergence Theorem in

(20.113) to show that the $h \to 0$ limit exists, and it is equal to

$$\lim_{h\to 0} \frac{I(x+he_i) - I(x)}{h} = \int_{\mathbb{R}^d} \lim_{h\to 0} \frac{F(x+he_i, y) - F(x,y)}{h} \, \mathrm{d}y = \int_{\mathbb{R}^d} \partial_i F(x,y) \, \mathrm{d}y. \tag{20.115}$$

So $\partial_i I$ exists! Finally, it's continuous, because for $E = \overline{B(x_0, r)} \subseteq U$ and $x \in E$,

$$\lim_{x\to x_0} \partial_i I(x) = \lim_{x\to x_0} \int_{\mathbb{R}^d} \partial_i F(x,y) \, \mathrm{d}y, \tag{20.116}$$

and the integrand is dominated by an integrable function, by hypothesis. Now, by Dominated Convergence Theorem again, we pass the $x \to x_0$ limit inside integral, and

$$\lim_{x\to x_0} \partial_i I(x) = \int_{\mathbb{R}^d} \lim_{x\to x_0} \partial_i F(x,y) \, \mathrm{d}y = \int_{\mathbb{R}^d} \partial_i F(x_0, y) \, \mathrm{d}y = \partial_i I(x_0), \tag{20.117}$$

since $\partial_i F(-, y)$ is continuous. Therefore, $I \in C^1(U)$. □

One can extend the same argument for C^∞ functions.

Claim 20.2: Let $F : U \times \mathbb{R}^d \to \mathbb{R}$ be a measurable function for $U \subseteq \mathbb{R}^n$ open, such that

1. $F(x, -)$ is L^1 for every $x \in U$;
2. $F(-, y)$ is C^∞ for a.e. $y \in \mathbb{R}^d$;
3. For every compact subset $E \subset U$ and every multi-index α, there exists $g_{E,\alpha} \in L^1(\mathbb{R}^d)$ such that
$$\sup_{x\in E} |\partial^\alpha F(x,y)| \le g_{E,\alpha}(y) \text{ for a.e. } y. \tag{20.118}$$

Then $I(x) = \int_{\mathbb{R}^d} F(x,y) \, \mathrm{d}y$ is C^∞ and $\partial^\alpha I(x) = \int_{\mathbb{R}^d} \partial^\alpha F(x,y) \, \mathrm{d}y$. (Here all the derivatives are with respect to x, and not y.)

Proof. This is just an application of Claim 20.1 and induction. We prove that $I \in C^k(U)$ for every k and $\partial^\alpha I(x) = \int_{\mathbb{R}^d} \partial^\alpha F(x) \, \mathrm{d}y$ for $|\alpha| \le k$, by induction on k. The base case $k = 1$ is just Claim 20.1. Now, suppose $I \in C^k(U)$. Let α be any index of order k.By hypothesis, $\partial^\alpha F(x, -)$ is L^1 for every x, $\partial^\alpha F(-, y)$ is C^1 for a.e. y, and for every i, there exists integrable $g_{E,\alpha+e_i}$ such that

$$\sup_{x\in E} |\partial_i \partial^\alpha F(x,y)| \le g_{E,\alpha+e_i}(x,y) \text{ for a.e. } y. \tag{20.119}$$

Therefore, by Claim 20.1, $\partial^\alpha I(x) = \int_{\mathbb{R}^d} \partial^\alpha F(x) \, \mathrm{d}y$ is C^1. This is true for every α with $|\alpha| = k$, so $I \in C^{k+1}$. So we are done by induction! □

Claim 20.3: Let $f \in L^p(\mathbb{R}^d)$ and $K : \mathbb{R}^d \times U \to \mathbb{R}$ for $U \subseteq \mathbb{R}^n$ open, and $I : \mathbb{R}^d \times U \to \mathbb{R}$

be defined as

$$I(x,t) := \int_{\mathbb{R}^d} f(y) K(x-y,t) \, \mathrm{d}y. \tag{20.120}$$

Suppose q be such that $\frac{1}{p} + \frac{1}{q} = 1$, and

1. $K(-,t) \in L^q$ for every $t \in U$;
2. $K \in C^\infty\left(\mathbb{R}^d \times U\right)$;
3. For every compact subset $E \subseteq \mathbb{R}^d \times U$ and every multi-index α, there exists $g_{E,\alpha} \in L^q$ such that

$$\sup_{(x,t)\in E} |\partial^\alpha K(x-y,t)| \leq g_{E,\alpha}(y) \text{ for a.e. } y. \tag{20.121}$$

Then $I \in C^\infty\left(\mathbb{R}^d \times U\right)$, and

$$\partial^\alpha I = \int_{\mathbb{R}^d} f(y) \partial^\alpha K(x-y,t) \, \mathrm{d}y. \tag{20.122}$$

(Here $\partial^\alpha K$ is understood to be derivatives with respect to (x,t).)

Proof. This is just an application of Hölder's inequality and Claim 20.2. Let $V = \mathbb{R}^d \times U$. Then I, on V is defined as $I(v) = \int_{\mathbb{R}^d} F(v,y) \, \mathrm{d}y$, where

$$F(v,y) = F((x,t),y) = f(y) K(x-y,t). \tag{20.123}$$

Now, we have to just check that F satisfies the hypotheses of Claim 20.2. For a fixed v, $F(v,-)$ is in L^1 since $y \mapsto f(y)$ is in L^p and $y \mapsto K(x-y,t)$ is in L^q (by hypothesis). Therefore, by Hölder's inequality, their product $y \mapsto F(v,y) = f(y) K(x-y,t)$ is in L^1. Furthermore, F is C^∞.

Also, for every multi-index α, and every compact $E \subseteq V$,

$$\sup_{v\in E} |\partial^\alpha F(v,y)| = |f(y)| \sup_{(x,t)\in E} |\partial^\alpha K(x-y,t)| \leq |f(y)| \, g_{E,\alpha}(y). \tag{20.124}$$

$f \in L^p$ and $g_{E,\alpha} \in L^q$, so by Hölder's inequality, their product is in L^1. Therefore, the hypotheses of Claim 20.2 hold, and hence, $I \in C^\infty(V)$ and

$$\partial^\alpha I = \int_{\mathbb{R}^d} \partial^\alpha F((x,t),y) \, \mathrm{d}y = \int_{\mathbb{R}^d} f(y) \partial^\alpha K(x-y,t) \, \mathrm{d}y \tag{20.125}$$

□

Now, we just need to check that our kernel satisfies the hypotheses of Claim 20.3. The kernel is

$$K(x,t) := K_t(x) = (4\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|x|^2}{4t}\right) \tag{20.126}$$

Let $E = \overline{B(0,R)} \times [t_1, t_2] \subseteq \mathbb{R}^d \times (0,\infty)$ be a compact set (any compact set is contained in a

compact set of this form). For any multi-index α and k, the partial derivative $\partial_t^k \partial_x^\alpha K$ is

$$\partial_t^k \partial_x^\alpha K(x-y,t) = \sum_{\text{finite linear combination}} (\text{const})\, t^{-\gamma} (x-y)^\beta \exp\left(-\frac{|x-y|^2}{4t}\right), \tag{20.127}$$

for some multi-index β and some positive γ. We bound each of these terms by an integrable function. For $t \geq t_1 > 0$, $t^{-\gamma} \leq t_1^{-\gamma}$. On the other hand,

$$\begin{aligned}
\left|(x-y)^\beta\right| = |x-y|^{|\beta|} &\leq 2^{|\beta|}\left(|x|^{|\beta|} + |y|^{|\beta|}\right) && (20.128)\\
&\leq 2^{|\beta|}\left(R^{|\beta|} + |y|^{|\beta|}\right) \leq 2^{|\beta|} \max\left\{1, R^{|\beta|}\right\}\left(1 + |y|^{|\beta|}\right) && (20.129)\\
&=: C_{\beta,R}\left(1 + |y|^{|\beta|}\right). && (20.130)
\end{aligned}$$

Now, by reverse triangle inequality,

$$|x-y| \geq |y| - |x| \geq |y| - R. \tag{20.131}$$

So we have

$$\exp\left(-\frac{|x-y|^2}{4t}\right) \leq \exp\left(-\frac{|y|^2}{4t_2}\right) \cdot \exp\left(\frac{R^2}{4t_2}\right). \tag{20.132}$$

After combining all these, we get

$$\left|t^{-\gamma}(x-y)^\beta \exp\left(-\frac{|x-y|^2}{4t}\right)\right| \leq C_{\beta,\gamma,t_1,t_2,R}\left(1 + |y|^{|\beta|}\right)\exp\left(-\frac{|y|^2}{4t_2}\right). \tag{20.133}$$

Summing up all these terms, we get

$$\left|\partial_t^k \partial_x^\alpha K(x-y,t)\right| \leq C\left(1 + |y|^M\right)\exp\left(-\frac{|y|^2}{4t_2}\right). \tag{20.134}$$

Claim 20.4: Let $P(x)$ be a polynomial on $\mathbb{R}$ and $a > 0$. Then $f(x) = P(x)e^{-ax^2}$ is in $L^p(\mathbb{R})$ for every $1 \leq p \leq \infty$.

Proof. It's easy to check that

$$\lim_{|x|\to\infty}\left|P(x)e^{-ax^2}\right| = 0. \tag{20.135}$$

So for $|x| > R$, $|f(x)| < 1$, and f is bounded on the compact interval $[-R, R]$. So f is bounded, i.e. $f \in L^\infty$. Now, consider the case $1 \leq p < \infty$.

For $|x| > 1$, suppose $P(x) = \sum_{i=0}^n c_i x^i$.

$$|P(x)| \leq \sum_{i=0}^n |c_i|\,|x|^i \leq C\,|x|^n, \quad \text{where } C = \sum_{i=0}^n |c_i|. \tag{20.136}$$

Therefore,

$$|P(x)|^p \leq C' |x|^{np} e^{-apx^2}. \tag{20.137}$$

Let $2m > np$. Then

$$e^{\frac{apx^2}{2}} = \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{apx^2}{2}\right)^j > \frac{1}{m!} \left(\frac{apx^2}{2}\right)^m = C'' x^{2m} > C'' x^{np}. \tag{20.138}$$

So we have

$$|x|^{np} e^{-apx^2} \leq |x|^{np} e^{-\frac{apx^2}{2}} e^{-\frac{apx^2}{2}} \leq \frac{1}{C''} e^{-\frac{apx^2}{2}}. \tag{20.139}$$

So

$$|P(x)|^p \leq C_1 e^{-\frac{apx^2}{2}}. \tag{20.140}$$

$e^{-\frac{apx^2}{2}}$ is integrable, so $|P(x)|^p$ is integrable on $\mathbb{R} - [-1, 1]$. On $[-1, 1]$, it's a bounded function. So $|P(x)|^p$ is integrable on $\mathbb{R}$. □

Now, $\left(1 + |y|^M\right) \exp\left(-\frac{|y|^2}{4t_2}\right)$ is in $L^p\left(\mathbb{R}^d\right)$, because this is a radial function, so its integral is the one dimensional r-integral (which is finite by Claim 20.4) times some constant. Therefore, $\left|\partial_t^k \partial_x^\alpha K\left(x - y, t\right)\right|$ is bounded by an L^q function. Therefore, by Claim 20.3, $u(x) = f * K_t(x)$ is smooth!

By Claim 20.3, if D is any differential operator (linear combination of ∂^α's),

$$Du = \int_{\mathbb{R}^d} f(y) DK\left(x - y, t\right) \, \mathrm{d}y \tag{20.141}$$

So it suffices to show that

$$\left(\partial_t - \sum_{i=1}^{d} \partial_i^2\right) K = 0. \tag{20.142}$$

Now,

$$\partial_t K\left(z, t\right) = \partial_t \left[(4\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|z|^2}{4t}\right)\right] \tag{20.143}$$

$$= (4\pi)^{-\frac{d}{2}} \left[-\frac{d}{2} t^{-\frac{d}{2}-1} + t^{-\frac{d}{2}} \frac{|z|^2}{4t^2}\right] \exp\left(-\frac{|z|^2}{4t}\right) \tag{20.144}$$

$$= \left[-\frac{d}{2t} + \frac{|z|^2}{4t^2}\right] K(z, t). \tag{20.145}$$

On the other hand,

$$\partial_i K\left(z, t\right) = \partial_i \left[(4\pi t)^{-\frac{d}{2}} \exp\left(-\frac{|z|^2}{4t}\right)\right] = -\frac{2z_i}{4t} K\left(z, t\right) = -\frac{z_i}{2t} K\left(z, t\right). \tag{20.146}$$

So

$$\partial_i^2 K\left(z, t\right) = \partial_i \left(-\frac{z_i}{t} K\left(z, t\right)\right) = \frac{z_i^2}{4t} K\left(z, t\right) - \frac{1}{2t} K\left(z, t\right). \tag{20.147}$$

Taking sum, we have

$$\sum_{i=1}^{d} \partial_i^2 K(z,t) = K(z,t)\left(\frac{|z|^2}{4t} - \frac{d}{2t}\right). \tag{20.148}$$

So $\partial_t K(z,t) = \sum_{i=1}^{d} \partial_i^2 K(z,t)$, and we are done!

Finally, K is a radial non-increasing kernel. So by Theorem 20.1,

$$\lim_{t\to 0^+} f * K_t(x) = f(x) \text{ for a.e. } x \in \mathbb{R}^d. \tag{20.149}$$

■

Exercise 20.4. The Poisson kernel $P : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$P(x) = \frac{C_d}{\left(1+|x|^2\right)^{\frac{d+1}{2}}}, \tag{20.150}$$

where $C_d = \Gamma\left(\frac{d+1}{2}\right) \Big/ \pi^{\frac{d+1}{2}}$. For $y > 0$ define $P_y(x) = y^{-d} P\left(\frac{x}{y}\right)$. Let $f \in L^p\left(\mathbb{R}^d\right)$, with $1 \le p \le \infty$. Prove that $u(x,y) = f * P_y(x)$ is harmonic in $\mathbb{R}^{d+1}_+ = \left\{(x,y);\ x \in \mathbb{R}^d,\ y > 0\right\}$ and that

$$\lim_{y\to 0^+} u(x,y) = f(x) \quad \text{a.e. } x \in \mathbb{R}^d. \tag{20.151}$$

Solution. Let's first compute ω_d.

Claim 20.5: $\omega_{d-1} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$.

Proof. Let's compute the d-dimsneional Gaussian integral $\int_{\mathbb{R}^d} e^{-|x|^2}\,\mathrm{d}x$ in two different ways. The integral factors into d one-dimensional Gaussians, each equal to $\sqrt{\pi}$. So by Fubini's theorem or Tonelli's theorem,

$$\int_{\mathbb{R}^d} e^{-|x|^2}\,\mathrm{d}x = \left(\int_{\mathbb{R}} e^{-t^2}\,\mathrm{d}t\right)^d = \pi^{\frac{d}{2}}. \tag{20.152}$$

Since the integrand is radial, we can also compute it in polar coordinates.

$$\int_{\mathbb{R}^d} e^{-|x|^2}\,\mathrm{d}x = \omega_{d-1}\int_0^\infty r^{d-1}e^{-r^2}\,\mathrm{d}r. \tag{20.153}$$

Substituting $t = r^2$, so $\mathrm{d}r = \frac{1}{2}t^{-\frac{1}{2}}\,\mathrm{d}t$, gives

$$\int_0^\infty r^{d-1}e^{-r^2}\,\mathrm{d}r = \frac{1}{2}\int_0^\infty t^{\frac{d}{2}-1}e^{-t}\,\mathrm{d}t = \frac{1}{2}\Gamma\left(\frac{d}{2}\right). \tag{20.154}$$

Comparing the two, we have

$$\pi^{\frac{d}{2}} = \omega_{d-1} \cdot \frac{1}{2}\Gamma\left(\frac{d}{2}\right) \implies \omega_{d-1} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}. \tag{20.155}$$

□

Then, we verify that $\int |P| = \int P = 1$. We convert the integral into polar coordinates:

$$\int_{\mathbb{R}^d} \frac{\mathrm{d}x}{\left(1+|x|^2\right)^{\frac{d+1}{2}}} = \omega_{d-1} \int_0^\infty \frac{r^{d-1}}{(1+r^2)^{\frac{d+1}{2}}}\,\mathrm{d}r, \tag{20.156}$$

where ω_d is the $(d-1)$-dimensional measure of S^{d-1}. Then we the substitution $s = \frac{r^2}{1+r^2}$, so

$$1 - s = \frac{1}{1+r^2}, \quad r = s^{\frac{1}{2}}(1-s)^{-\frac{1}{2}}, \quad \mathrm{d}r = \frac{1}{2}s^{-\frac{1}{2}}(1-s)^{-\frac{3}{2}}\,\mathrm{d}s. \tag{20.157}$$

So the integral becomes

$$\int_0^\infty \frac{r^{d-1}}{(1+r^2)^{\frac{d+1}{2}}}\,\mathrm{d}r = \frac{1}{2}\int_0^1 s^{\frac{d}{2}-1}(1-s)^{\frac{1}{2}-1}\,\mathrm{d}s = \frac{1}{2}B\left(\frac{d}{2},\frac{1}{2}\right) \tag{20.158}$$

$$= \frac{\Gamma\left(\frac{d}{2}\right)\Gamma\left(\frac{1}{2}\right)}{2\,\Gamma\left(\frac{d+1}{2}\right)}. \tag{20.159}$$

Using $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ and multiplying by $\omega_{d-1} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}$, we have

$$\int_{\mathbb{R}^d} \frac{\mathrm{d}x}{\left(1+|x|^2\right)^{\frac{d+1}{2}}} = \frac{2\pi^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \cdot \frac{\Gamma\left(\frac{d}{2}\right)\sqrt{\pi}}{2\,\Gamma\left(\frac{d+1}{2}\right)} = \frac{\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)} = C_d^{-1}. \tag{20.160}$$

So $\int P = 1$.

Claim 20.6: Let (Ω, Σ, μ) be a measure space and $1 \le a < b \le \infty$. Then $L^a(\mu) \cap L^b(\mu) \subseteq L^p(\mu)$ for every $p \in [a, b]$.

Proof. Let $f \in L^a(\mu) \cap L^b(\mu)$. If $p = a$ or $p = b$, there is nothing to prove, so assume $a < p < b$. Partition Ω into the measurable sets

$$E = \{x \in \Omega : |f(x)| \le 1\} \quad \text{and} \quad E^c = \{x \in \Omega : |f(x)| > 1\}. \tag{20.161}$$

On E, since $p > a$, we have $|f(x)|^p \le |f(x)|^a$. Hence

$$\int_E |f|^p\,\mathrm{d}\mu \le \int_E |f|^a\,\mathrm{d}\mu \le \|f\|_{L^a(\mu)}^a. \tag{20.162}$$

Case 1: $b < \infty$**.** On E^c, since $p < b$, we have $|f(x)|^p \leq |f(x)|^b$. Thus,

$$\int_\Omega |f|^p \, \mathrm{d}\mu = \int_E |f|^p \, \mathrm{d}\mu + \int_{E^c} |f|^p \, \mathrm{d}\mu \tag{20.163}$$

$$\leq \int_E |f|^a \, \mathrm{d}\mu + \int_{E^c} |f|^b \, \mathrm{d}\mu \tag{20.164}$$

$$\leq \|f\|_{L^a(\mu)}^a + \|f\|_{L^b(\mu)}^b < \infty. \tag{20.165}$$

Case 2: $b = \infty$**.** First, note that $\mu(E^c)$ is finite, since $|f| > 1$ on E^c:

$$\mu(E^c) = \int_{E^c} 1 \, \mathrm{d}\mu \leq \int_{E^c} |f|^a \, \mathrm{d}\mu \leq \|f\|_{L^a(\mu)}^a < \infty. \tag{20.166}$$

Since $|f(x)| \leq \|f\|_{L^\infty(\mu)}$ for μ-almost every x, we get

$$\int_\Omega |f|^p \, \mathrm{d}\mu = \int_E |f|^p \, \mathrm{d}\mu + \int_{E^c} |f|^p \, \mathrm{d}\mu \tag{20.167}$$

$$\leq \|f\|_{L^a(\mu)}^a + \|f\|_{L^\infty(\mu)}^p \, \mu(E^c) < \infty. \tag{20.168}$$

Hence, $f \in L^p(\mu)$ in all cases. □

Claim 20.7: $K(-, y) := P_y$ is in L^p for every $1 \leq p \leq \infty$.

Proof. We've just shown it's in L^1. It's in L^∞ since it's bounded by $y^{-d} C_d$, as

$$\frac{1}{\left(1 + |x|^2\right)^{\frac{d+1}{2}}} \leq 1. \tag{20.169}$$

Then by Claim 20.6, we are done! □

We show that u is smooth using Claim 20.3. We've already shown that $K(-, y)$ is in L^q for every q.

$$K(x, y) := P_y(x) = \frac{1}{y^d} P\left(\frac{x}{y}\right) = \frac{1}{y^d} \frac{C_d}{\left(1 + \frac{|x|^2}{y^2}\right)^{\frac{d+1}{2}}} = \frac{C_d y}{\left(y^2 + |x|^2\right)^{\frac{d+1}{2}}}. \tag{20.170}$$

is indeed smooth since it's the quotient of two strictly positive polnomials.

Let $E = \overline{B(0, R)} \times [y_1, y_2] \subseteq \mathbb{R}^d \times (0, \infty)$ be a compact set (any compact set is contained in a compact set of this form). For any multi-index α, the partial derivative $\partial^\alpha K$ is

$$\partial^\alpha K(w, y) = \sum_{\text{finite linear combination}} (\text{const}) \frac{w^\beta y^m}{\left(|w|^2 + y^2\right)^{\frac{d+1}{2} + j}}, \tag{20.171}$$

where β is a multi-index, $m, j \in \mathbb{Z}_{\geq 0}$, and $|\beta| + m = 2j + 1 - |\alpha|$. Using

$$\left|w^{\beta}\right| \leq \left(|w|^2 + y^2\right)^{\frac{|\beta|}{2}} \text{ and } y^m \leq \left(|w|^2 + y^2\right)^{\frac{m}{2}}, \tag{20.172}$$

we obtain the uniform bound for all $|\alpha| \geq 1$,

$$|\partial^{\alpha} K(w, y)| \leq \frac{C_{d,\alpha}}{\left(|w|^2 + y^2\right)^{\frac{d+|\alpha|}{2}}}. \tag{20.173}$$

Putting $w = x - z$, and noting that $y \geq y_1$, we have

$$|\partial^{\alpha} K(x - z, y)| \leq \frac{C_{d,\alpha}}{\left(|x - z|^2 + y_1^2\right)^{\frac{d+|\alpha|}{2}}}. \tag{20.174}$$

For $|z| > 2R$,

$$|x - z| \geq |z| - |x| \geq |z| - \frac{1}{2}|z| = \frac{1}{2}|z|. \tag{20.175}$$

So we have for $|z| > 2R$,

$$|\partial^{\alpha} K(x - z, y)| \leq \frac{C_{d,\alpha}}{\left(\frac{1}{4}|z|^2 + y_1^2\right)^{\frac{d+|\alpha|}{2}}} \leq C' \frac{1}{\left(|z|^2 + 1\right)^{\frac{d+|\alpha|}{2}}} \tag{20.176}$$

where $C' = \frac{C_{d,\alpha}}{\min\left\{\frac{1}{4}, y_1^2\right\}}$. Now, on $|z| \leq 2R$, we can bound $|\partial^{\alpha} K(x - z, y)|$ by the constant $C_{d,\alpha}\ y_1^{-d-|\alpha|}$. So we have

$$\sup_{(x,y)\in E} |\partial^{\alpha} K(x - z, y)| \leq \begin{cases} C_{d,\alpha}\ y_1^{-d-|\alpha|} & \text{if } |z| \leq 2R, \\ C' \left(|z|^2 + 1\right)^{-\frac{d+|\alpha|}{2}} & \text{if } |z| > 2R. \end{cases} \tag{20.177}$$

We have to show that this bounding function is in L^q.

Claim 20.8: $\left(|z|^2 + 1\right)^{-\frac{d+|\alpha|}{2}} \in L^p\left(\mathbb{R}^d - \overline{B(0, 2R)}\right)$ for every $1 \leq p \leq \infty$.

Proof. It's in L^{∞} since it's bounded by $\left(R^2 + 1\right)^{-\frac{d+|\alpha|}{2}}$. Now, let's verify it's in L^1.

$$\begin{aligned} \int_{\overline{B(0,2R)}^c} \left(|z|^2 + 1\right)^{-\frac{d+|\alpha|}{2}} &= \omega_d \int_{2R}^{\infty} \left(r^2 + 1\right)^{-\frac{d+|\alpha|}{2}}\ r^{d-1}\, \mathrm{d}r \\ &\leq \omega_d \int_{2R}^{\infty} \left(r^2\right)^{-\frac{d+|\alpha|}{2}}\ r^{d-1}\, \mathrm{d}r \\ &= \omega_d \int_{2R}^{\infty} r^{-1-\alpha}\, \mathrm{d}r = \omega_d \left[\frac{r^{-\alpha}}{-\alpha}\right]_{2R}^{\infty} \\ &= \omega_d \frac{(2R)^{-\alpha}}{\alpha} < \infty. \end{aligned} \tag{20.178}$$

So it's in L^1. Therefore, it's in all L^p, by Claim 20.6. □

So, all the hypotheses of Claim 20.3 are satisfied. Therefore, $u : \mathbb{R}^d \times (0,\infty)$ is smooth, and we can differentiate under the integral sign. By Claim 20.3, if D is any differential operator (linear combination of ∂^α's),

$$Du = \int_{\mathbb{R}^d} f(z) DK(x-z,y)\, \mathrm{d}z. \tag{20.179}$$

So in order to prove the harmonicity of u, i.e. $\nabla^2 u = 0$, it suffices to show the harmonicity of K. Let $r^2 = y^2 + |x|^2$.

$$\begin{aligned}
\frac{\partial K}{\partial x_i} &= -\frac{d+1}{2} C_d y \left(y^2+|x|^2\right)^{-\frac{d+3}{2}} 2x_i = -(d+1) C_d y x_i r^{-d-3}. \\
\frac{\partial^2 K}{\partial x_i^2} &= -(d+1) C_d y r^{-d-3} - (d+1) C_d y x_i \left(-\frac{d+3}{2}\right)\left(y^2+|x|^2\right)^{-\frac{d+5}{2}} 2x_i \\
&= -(d+1) C_d y r^{-d-3} + (d+1)(d+3) C_d y x_i^2 r^{-d-5}. \\
&= (d+1) C_d y \left[-r^{-d-3} + (d+3) x_i^2 r^{-d-5}\right] \\
\frac{\partial K}{\partial y} &= C_d r^{-d-1} - \frac{d+1}{2} C_d y \left(y^2+|x|^2\right)^{-\frac{d+3}{2}} 2r = C_d r^{-d-1} - (d+1) C_d y^2 r^{-d-3}. \\
\frac{\partial^2 K}{\partial y^2} &= -(d+1) y C_d r^{-d-3} - 2y(d+1) C_d r^{-d-3} + (d+1)\frac{d+3}{2} C_d y^3 r^{-d-5} 2y \\
&= (d+1) C_d y \left[-3r^{-d-3} + (d+3) y^2 r^{-d-5}\right].
\end{aligned}$$

Summing up everything, we get

$$\begin{aligned}
\nabla^2 K &= (d+1) C_d y \left[\sum_{i=1}^d \left(-r^{-d-3} + (d+3) x_i^2 r^{-d-5}\right) - 3r^{-d-3} + (d+3) y^2 r^{-d-5}\right] \\
&= (d+1) C_d y r^{-d-3} \left[-d - 3 + (d+3) r^{-2}\left(\sum_{i=1}^d x_i^2 + y^2\right)\right] \\
&= 0.
\end{aligned} \tag{20.180}$$

So $\nabla^2 K = 0$, and hence, $\nabla^2 u = 0$.

Finally, P is a radially decreasing kernel, and $A := \int P = 1$. So, by Theorem 20.1 (Approximations of the identity: pointwise convergence),

$$\lim_{y\to 0^+} u(x,y) = \lim_{y\to 0^+} P_y * f(x) = Af(x) = f(x) \text{ for a.e. } x \in \mathbb{R}^d. \tag{20.181}$$

■

Exercise 20.5. For $y > 0$ define

$$\varphi_y(x) = \frac{1}{\pi} \sin\left(\frac{\pi x y}{x^2+y^2}\right) \frac{y}{x^2+y^2}; \tag{20.182}$$

$$\varphi_y * f(x) = \int_{-\infty}^{\infty} \varphi_y(x-z) f(z)\, \mathrm{d}z, \qquad f \in L^2(\mathbb{R}). \tag{20.183}$$

(i) Show that $K(x,y) = \sqrt{y}\,(\varphi_y * f)(x)$ is a bounded function of x and y with $(x,y) \in \mathbb{R} \times (0,\infty)$.

(ii) For each $y > 0$, show that $\varphi_y * f \in L^2(\mathbb{R})$. Show also that $\varphi_y * f$ converges in $L^2(\mathbb{R})$ as $y \to 0$ and find this limit.

Solution. (i) First note that if $\Phi = \varphi_1$, then first notice that $\Phi \in L^1$. Indeed, since $|\sin\theta| \le 1$, so

$$\int_{\mathbb{R}} |\Phi(x)|\,\mathrm{d}x \le \frac{1}{\pi}\int_{\mathbb{R}} \frac{1}{x^2+1}\,\mathrm{d}x = \frac{2}{\pi}\int_0^\infty \frac{1}{x^2+1}\,\mathrm{d}x = \frac{2}{\pi}\frac{\pi}{2} = 1. \tag{20.184}$$

Φ is an odd function so in integrates to 0. In other words,

$$\int_{\mathbb{R}} \Phi(x)\,\mathrm{d}x = \frac{1}{\pi}\int_{\mathbb{R}} \sin\left(\frac{\pi x}{x^2+1}\right)\frac{1}{x^2+1}\,\mathrm{d}x, \tag{20.185}$$

and if we do a change of variables $u = -x$

$$\int_{\mathbb{R}} \Phi(x)\,\mathrm{d}x = -\frac{1}{\pi}\int_{\infty}^{-\infty} -\sin\left(\frac{\pi u}{u^2+1}\right)\frac{1}{x^2+1}\,\mathrm{d}u = -\int_{\mathbb{R}} \Phi(u)\,\mathrm{d}u. \tag{20.186}$$

So we have

$$\int_{\mathbb{R}} \Phi(x)\,\mathrm{d}x = 0. \tag{20.187}$$

Now, we claim that $\Phi \in L^2$. Indeed, since $|\sin\theta| \le 1$, so

$$\int_{\mathbb{R}} |\Phi(x)|^2\,\mathrm{d}x \le \frac{1}{\pi^2}\int_{\mathbb{R}} \frac{1}{(x^2+1)^2}\,\mathrm{d}x = \frac{2}{\pi^2}\int_0^\infty \frac{1}{(x^2+1)^2}\,\mathrm{d}x. \tag{20.188}$$

Putting $x = \tan\theta$, we have $\mathrm{d}\theta = \frac{\mathrm{d}x}{1+x^2}$. So

$$\int_0^\infty \frac{1}{(x^2+1)^2}\,\mathrm{d}x = \int_0^{\frac{\pi}{2}} \frac{1}{\tan^2\theta+1}\,\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \cos^2\theta\,\mathrm{d}\theta \tag{20.189}$$

$$= \frac{1}{2}\int_0^{\frac{\pi}{2}} (1+\cos 2\theta)\,\mathrm{d}\theta = \frac{\pi}{4} + \frac{1}{4}\Big[\sin 2\theta\Big]_0^{\frac{\pi}{2}} \tag{20.190}$$

$$= \frac{\pi}{4}. \tag{20.191}$$

So

$$\int_{\mathbb{R}} |\Phi(x)|^2\,\mathrm{d}x \le \frac{2}{\pi^2}\frac{\pi}{4} = \frac{1}{2\pi}. \tag{20.192}$$

Now, in the sense of Φ_ε,

$$\Phi_y(x) = \frac{1}{y}\Phi\left(\frac{x}{y}\right) = \frac{1}{y}\frac{1}{\pi}\sin\left(\frac{\pi(x/y)}{(x/y)^2+1}\right)\frac{1}{(x/y)^2+1} \tag{20.193}$$

$$= \frac{1}{y}\frac{1}{\pi}\sin\left(\frac{\pi xy}{x^2+y^2}\right)\frac{y^2}{x^2+y^2} \tag{20.194}$$

$$= \frac{1}{\pi}\sin\left(\frac{\pi xy}{x^2+y^2}\right)\frac{y}{x^2+y^2} = \varphi_y(x). \tag{20.195}$$

Now,

$$\|\Phi_y\|_2^2 = \int_{\mathbb{R}} |\Phi_y(x)|^2 \, \mathrm{d}x = \frac{1}{y} \int_{\mathbb{R}} \frac{1}{y} \left| \Phi\left(\frac{x}{y}\right) \right|^2 \mathrm{d}x \tag{20.196}$$

$$= \frac{1}{y} \int_{\mathbb{R}} \left(|\Phi|_y \right)^2 (x) \mathrm{d}x = \frac{1}{y} \int_{\mathbb{R}} |\Phi|^2 (x) \mathrm{d}x \tag{20.197}$$

$$= \frac{1}{y} \|\Phi\|_2^2 . \tag{20.198}$$

Now,

$$|K(x,y)| = \left| \int_{\mathbb{R}} \sqrt{y} \varphi_y (x-z) f(z) \, \mathrm{d}z \right| \leq \|\sqrt{y}\varphi_y\|_2 \|f\|_2 , \tag{20.199}$$

by Hölder's inequality. By (20.198),

$$\|\sqrt{y}\varphi_y\|_2 = \|\sqrt{y}\phi_y\|_2 = \|\Phi\|_2 . \tag{20.200}$$

So we have

$$|K(x,y)| \leq \|\Phi\|_2 \|f\|_2 , \tag{20.201}$$

both of which are finite numbers. So K is bounded on $\mathbb{R} \times (0, \infty)$.

(ii) Young's inequality says that if $1 + \frac{1}{r} = \frac{1}{p} + \frac{1}{q}$, and $f \in L^p, g \in L^q$, then

$$f * g \in L^r. \tag{20.202}$$

In our case, $\Phi_y \in L^1$ and $f \in L^2$, and $1 + \frac{1}{2} = \frac{1}{1} + \frac{1}{2}$, so

$$\Phi_y * f \in L^2. \tag{20.203}$$

Now, by Approximations of the identity: L^p-convergence, let $A := \int \Phi = 0$,

$$\lim_{y \to 0^+} \|\Phi_y * f - Af\|_2 = 0. \tag{20.204}$$

So $\Phi_y * f \to 0$ in L^2 as $y \to 0^+$.

■

Exercise 20.6. Let A and B be measurable subsets of $\mathbb{R}$, such that $m(A) > 0$ and $m(B) > 0$. Show that the set

$$A + B = \{x + y; \ x \in A, \ y \in B\} \tag{20.205}$$

contains an interval.

Solution. WLOG, we can assume $0 < m(A), m(B) < \infty$, because if $m(A) = \infty$, then A contains a subset A' of finite measure. If we show that $A' + B$ contains an interval, then that would prove $A + B$ contains an interval.

First, notice that $\chi_A, \chi_B \in L^2$, since

$$\int_{\mathbb{R}} \chi_A^2 = m(A) < \infty, \quad \int_{\mathbb{R}} \chi_B^2 = m(B) < \infty. \tag{20.206}$$

Therefore, by Proposition 19.4, the convolution $f := \chi_A * \chi_B$ is uniformly continuous.

$$f(x) = \int_{\mathbb{R}} \chi_A(y)\, \chi_B(x-y)\ \mathrm{d}y. \tag{20.207}$$

Let's now compute $\int f$:

$$\int_{\mathbb{R}} f(x)\ \mathrm{d}x = \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \chi_A(y)\, \chi_B(x-y)\ \mathrm{d}y \right] \mathrm{d}x \tag{20.208}$$

By Tonelli's theorem,

$$\int_{\mathbb{R}} f(x)\ \mathrm{d}x = \int_{\mathbb{R}} \chi_A(y) \left[\int_{\mathbb{R}} \chi_B(x-y)\ \mathrm{d}x \right] \mathrm{d}y \tag{20.209}$$

$$= \int_{\mathbb{R}} \chi_A(y)\, [m(B)]\, \mathrm{d}y = m(A)m(B) > 0. \tag{20.210}$$

$f \geq 0$ with $\int_{\mathbb{R}} f > 0$. So we have $f(x_0) = 0$ for some x_0. Since f is continuous, there exists some interval $I := (x_0 - \varepsilon, x_0 + \varepsilon)$ such that $f(x) > 0$ for $x \in I$. Now, note that for $x \in I$,

$$f(x) = \int_{\mathbb{R}} \chi_A(y)\, \chi_B(x-y)\ \mathrm{d}y > 0, \tag{20.211}$$

so the integrand is positive for some y. In other words, there exists some $y \in A$ such that $x - y \in B$. This proves that $x = y + (x-y) \in A + B$. So $I \subseteq A + B$. ■

Exercise 20.7. Let $I = [-M, M] \subseteq \mathbb{R}$ and let $h : \mathbb{R} \to \mathbb{R}$ be a continuous function with support in I. Define, for $z \in \mathbb{C}$,

$$g(z) = \int_I \frac{1}{t-z}\, h(t)\ \mathrm{d}t. \tag{20.212}$$

(i) Prove that g is analytic in $\mathbb{C} \setminus I$.

(ii) Find the limit $\lim_{|z|\to\infty} z\, g(z)$.

(iii) For $\varepsilon > 0$ and $\sigma \in \mathbb{R}$, find the limit

$$\lim_{\varepsilon \to 0} \left[g(\sigma + i\varepsilon) - g(\sigma - i\varepsilon) \right]. \tag{20.213}$$

Solution. (i) We claim that g is differentiable everywhere in $\mathbb{C} - I$ with $g'(z) = \int_I \frac{1}{(t-z)^2} h(t)\, \mathrm{d}t$. Let $z_0 \in \mathbb{C} - I$. Since I is a compact interval, $d := \operatorname{dist}(z_0, I) > 0$. Let $w \in \mathbb{C} - I$ with

$|w| < \frac{d}{2}$. Then

$$\frac{g(z_0+w)-g(z_0)}{w} = \frac{1}{w}\int_I \left(\frac{1}{t-z_0-w} - \frac{1}{t-z_0}\right) h(t)\,\mathrm{d}t \tag{20.214}$$

$$= \frac{1}{w}\int_I \frac{t-z_0-t+z_0+w}{(t-z_0-w)(t-z_0)} h(t)\,\mathrm{d}t \tag{20.215}$$

$$= \int_I \frac{1}{(t-z_0-w)(t-z_0)} h(t)\,\mathrm{d}t \tag{20.216}$$

Then

$$\begin{aligned} D &:= \left|\frac{g(z_0+w)-g(z_0)}{w} - \int_I \frac{1}{(t-z_0)^2} h(t)\,\mathrm{d}t\right| \\ &= \left|\int_I \left(\frac{1}{(t-z_0-w)(t-z_0)} - \frac{1}{(t-z_0)^2}\right) h(t)\,\mathrm{d}t\right| \\ &= \left|\int_I \left(\frac{w}{(t-z_0-w)(t-z_0)^2}\right) h(t)\,\mathrm{d}t\right| \\ &\le |w| \int_I \left(\frac{1}{\left|(t-z_0-w)(t-z_0)^2\right|}\right) |h(t)|\,\mathrm{d}t. \end{aligned} \tag{20.217}$$

Now, $|t-z_0| \ge d$, and

$$|t-(z_0+w)| \ge |t-z_0| - |w| \ge d - \frac{d}{2} = \frac{d}{2}. \tag{20.218}$$

Therefore,

$$D \le |w| \int_I \frac{1}{\frac{d}{2}\cdot d^2} |h(t)|\,\mathrm{d}t = \frac{2|w|}{d^3}\int_I |h(t)|\,\mathrm{d}t. \tag{20.219}$$

h is a continuous function supported on the compact interval I, so $\int_I |h(t)|\,\mathrm{d}t < \infty$. Hence, $D \to 0$ as $|w| \to 0$. In other words,

$$\lim_{w\to 0} \frac{g(z_0+w)-g(z_0)}{w} = \int_I \frac{1}{(t-z_0)^2} h(t)\,\mathrm{d}t. \tag{20.220}$$

Since $z_0 \in \mathbb{C} - I$ was arbitrary, g is complex differentiable (i.e. analytic) on $\mathbb{C} - I$.

(ii) We'll show that $zg(z) \to -\int_I h(t)\,\mathrm{d}t$ as $|z| \to \infty$.

$$zg(z) + \int_I h(t)\,\mathrm{d}t = \int_I \left(\frac{z}{t-z} + 1\right) h(t)\,\mathrm{d}t = \int_I \frac{t}{t-z} h(t)\,\mathrm{d}t. \tag{20.221}$$

We have $|t| \le M$, and for $|z| > M$,

$$|t-z| \ge |z| - |t| \ge |z| - M. \tag{20.222}$$

Therefore,

$$\left|zg(z) + \int_I h(t)\,\mathrm{d}t\right| \le \int_I \left|\frac{t}{t-z}\right| |h(t)|\,\mathrm{d}t \le \frac{|M|}{|z|-M}\int_I |h(t)|\,\mathrm{d}t. \tag{20.223}$$

h is a continuous function supported on the compact interval I, so $\int_I |h(t)|\,\mathrm{d}t < \infty$. So $|z| \to \infty$, the RHS of (20.223) approaches 0. Therefore,

$$\lim_{|z|\to\infty} zg(z) = -\int_I h(t)\,\mathrm{d}t. \tag{20.224}$$

(iii)

$$g(\sigma+i\varepsilon) - g(\sigma-i\varepsilon) = \int_I \left(\frac{1}{t-\sigma-i\varepsilon} - \frac{1}{t-\sigma+i\varepsilon}\right) h(t)\,\mathrm{d}t \tag{20.225}$$

$$= \int_I \frac{2i\varepsilon}{(t-\sigma-i\varepsilon)(t-\sigma+i\varepsilon)} h(t)\,\mathrm{d}t \tag{20.226}$$

$$= 2i\int_I \frac{\varepsilon}{(t-\sigma)^2+\varepsilon^2}\, h(t)\,\mathrm{d}t. \tag{20.227}$$

Let $\Phi : \mathbb{R} \to \mathbb{R}$ be defined as

$$\Phi(x) = \frac{1}{x^2+1}. \tag{20.228}$$

Then $\Phi \in L^1$, since

$$\int_{\mathbb{R}} |\Phi(x)|\,\mathrm{d}x = \int_{\mathbb{R}} \Phi(x)\,\mathrm{d}x = 2\int_0^\infty \frac{1}{x^2+1}\,\mathrm{d}x = 2\,\frac{\pi}{2} = \pi. \tag{20.229}$$

Then

$$\Phi_\varepsilon(x) = \frac{1}{\varepsilon}\Phi\left(\frac{x}{\varepsilon}\right) = \frac{1}{\varepsilon}\frac{1}{\left(\frac{x}{\varepsilon}\right)^2+1} = \frac{1}{\varepsilon}\frac{\varepsilon^2}{x^2+\varepsilon^2} = \frac{\varepsilon}{x^2+\varepsilon^2}. \tag{20.230}$$

So we have from (20.227),

$$g(\sigma+i\varepsilon) - g(\sigma-i\varepsilon) = 2i\int_I \Phi_\varepsilon(t-\sigma)\,h(t)\,\mathrm{d}t \tag{20.231}$$

$$= 2i\int_{\mathbb{R}} \Phi_\varepsilon(t-\sigma)\,h(t)\,\mathrm{d}t \tag{20.232}$$

$$= 2i\,(\Phi_\varepsilon * h)(\sigma). \tag{20.233}$$

Here we replaced the integral over I to an integral over $\mathbb{R}$ since $\operatorname{supp} h \subseteq I$, so $h = 0$ outside of I and that part doesn't contribute to the integral.

Now, let $A := \int_{\mathbb{R}} \Phi(x)\,\mathrm{d}x = \pi$. h is a compactly supported continuous functon, so it's bounded and uniformly continuous. Therefore, by Theorem 19.7(ii) (Approximations of the identity: L^p-convergence), $\Phi_\varepsilon * h \to Ah$ in L^∞ norm. In other words,

$$\lim_{\varepsilon\to 0} \operatorname*{ess\,sup}_{x\in\mathbb{R}} |\Phi_\varepsilon * Ah(x) - h(x)| = 0. \tag{20.234}$$

Claim 20.9: If $f : \mathbb{R} \to \mathbb{R}$ is bounded continuous, then $\sup_{x\in\mathbb{R}} |f(x)| = \operatorname{ess\,sup}_{x\in\mathbb{R}} |f(x)|$.

Proof. sup $\geq$ ess sup is obvious. Suppose $|f(x_0)| > M$ for some x_0. Since f is

continuous, there's an interval $I := (x_0 - \delta, x_0 + \delta)$ such that $|f| > M$ on I. Therefore, $\{|f| > M\}$ can't have measure 0. Thus,

$$\operatorname*{ess\,sup}_{x\in\mathbb{R}} |f(x)| \geq M. \tag{20.235}$$

Taking $M = \sup |f| - \varepsilon$, we have

$$\operatorname*{ess\,sup}_{x\in\mathbb{R}} |f(x)| \geq \sup_{x\in\mathbb{R}} |f(x)| - \varepsilon. \tag{20.236}$$

This is true for arbitrarily small ε, so we have ess sup $\geq$ sup. □

Now, $\Phi_\varepsilon \in L^1$ and $h \in L^\infty$, so by Proposition 19.4, $\Phi_\varepsilon * h$ is bounded and uniformly continuous. Therefore, (20.234) translates to (in light of Claim 20.9),

$$\lim_{\varepsilon\to 0} \sup_{x\in\mathbb{R}} |\Phi_\varepsilon * h\,(x) - Ah\,(x)| = 0. \tag{20.237}$$

Therefore, we have the uniform convergence

$$\lim_{\varepsilon\to 0} \Phi_\varepsilon * h\,(x) = Ah\,(x)\,. \tag{20.238}$$

In other words,

$$\lim_{\varepsilon\to 0} g\,(\sigma + i\varepsilon) - g\,(\sigma - i\varepsilon) = 2\pi i\ h\,(\sigma)\,. \tag{20.239}$$

■

§21. 9 November 2026

§21.i. Recap: duality between decay and regularity

Recall that, for $f \in L^1\left(\mathbb{R}^d\right)$, the Fourier transform is

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i x\cdot\xi} \mathrm{d}x, \tag{21.1}$$

where $x \cdot \xi$ is the inner product in $\mathbb{R}^d$. This is a well-defined object, because $\left|e^{-2\pi i x\cdot\xi}\right| = 1$ and $f \in L^1$; as $|\xi|$ grows, the kernel $e^{-2\pi i x\cdot\xi}$ oscillates faster and faster. We have seen the basic properties in Proposition 20.6: translations become modulations, convolutions become products, and multiplication by polynomials becomes differentiation (and conversely).

Remark 21.1 (Duality: decay versus regularity)**.** The properties (v) and (vi) of Proposition 20.6 express a duality between *decay* on one side and *regularity* on the other side.

1. If f decays fast, in the sense that $x^\alpha f \in L^1\left(\mathbb{R}^d\right)$ for $|\alpha| \le k$ (the function integrates against polynomials), then $\widehat{f} \in C^k\left(\mathbb{R}^d\right)$.
2. If f is regular, in the sense that $f \in C^k$ and $\partial^\alpha f \in L^1$ (and $\partial^\alpha f \in C_0$ for $|\alpha| \le k-1$), then by (vi) and (i), for $|\alpha| \le k$,

$$\left|\xi^\alpha \widehat{f}(\xi)\right| = \frac{1}{(2\pi)^{|\alpha|}} \left|\left(\partial^\alpha f\right)^\wedge(\xi)\right| \le \frac{1}{(2\pi)^{|\alpha|}} \left\|\partial^\alpha f\right\|_{L^1}, \tag{21.2}$$

   so $\widehat{f}$ decays at infinity at a polynomial rate.

Decay on one side is regularity on the other side.

Remark 21.2 (Fractional derivatives)**.** This gives a way to define *fractional* derivatives. Everybody knows what it means for a function $f : \mathbb{R} \to \mathbb{R}$ to have one derivative, or two, but what does it mean to have "half" a derivative? One can make sense of this pointwise in a few ways, but globally the way to do it is via the Fourier transform. For instance, for the Laplacian $\Delta = \sum_{j=1}^d \partial_j^2$, property (vi) of Proposition 20.6 gives, since $(2\pi i\xi_j)^2 = -4\pi^2\xi_j^2$,

$$(\Delta f)^\wedge(\xi) = -4\pi^2 |\xi|^2 \widehat{f}(\xi), \tag{21.3}$$

i.e. the operator $-\Delta$ becomes multiplication by $4\pi^2|\xi|^2$ on the Fourier side. So, if we want to define a power $(-\Delta)^s$ of the Laplacian for $s > 0$ (for instance $s = \frac{1}{2}$), it is natural to *define* it by: take the Fourier transform, multiply by $\left(4\pi^2|\xi|^2\right)^s$, and go back with the inverse Fourier transform (to be studied below),

$$(-\Delta)^s f := \left(\left(4\pi^2|\xi|^2\right)^s \widehat{f}\right)^\vee. \tag{21.4}$$

This is one way of making sense of fractional derivatives.

Remark 21.3 (Algebra homomorphism)**.** The space $L^1(\mathbb{R}^d)$ with the usual sum and the product given by the convolution is an *algebra*: a vector space with a bilinear and associative multiplication (here also commutative; Proposition 19.1). Similarly, $C_0(\mathbb{R}^d)$ with the pointwise product is an associative algebra. By the Riemann–Lebesgue lemma below, $f \mapsto \widehat{f}$ maps L^1 into C_0; it is linear, and by property (iv) it takes convolutions to products:

$$(f * g)^\wedge = \widehat{f}\,\widehat{g}. \tag{21.5}$$

So the Fourier transform is a homomorphism of (associative) algebras from $(L^1, *)$ to $(C_0, \cdot)$.

§21.ii. Examples

We have defined an object, so it is natural to compute it on some examples: characteristic functions of intervals and cubes, and the Gaussian. We write

$$\operatorname{sinc} x = \frac{\sin(\pi x)}{\pi x} \quad \text{for } x \neq 0, \qquad \operatorname{sinc} 0 = 1. \tag{21.6}$$

Example 21.1 (Intervals and cubes)

Let $d = 1$ and $f = \chi_{[-\frac{1}{2},\frac{1}{2}]}$ (we put the factor $\frac{1}{2}$ so that the interval has length 1). For $\xi \neq 0$,

$$\widehat{f}(\xi) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-2\pi i x\xi} \mathrm{d}x = \left[\frac{e^{-2\pi i x\xi}}{-2\pi i\xi}\right]_{x=-\frac{1}{2}}^{x=\frac{1}{2}} = \frac{e^{-\pi i\xi} - e^{\pi i\xi}}{-2\pi i\xi}$$
$$= \frac{\sin(\pi\xi)}{\pi\xi} = \operatorname{sinc}\xi,$$

and for $\xi = 0$ we have $\widehat{f}(0) = \int f = 1$ (in general, $\widehat{f}(0)$ is always the integral of f); so $\widehat{f} = \operatorname{sinc}$ is continuous at 0, and $\widehat{f}(\xi) \to 0$ as $|\xi| \to \infty$.

Dilations. Let $a > 0$ and $\chi_{[-a,a]}(x) = f\left(\frac{x}{2a}\right) = f \circ T(x)$ with $Tx = \frac{x}{2a}$. By property (iii) of Proposition 20.6 ($|\det T|^{-1} = 2a$ and $(T^*)^{-1}\xi = 2a\xi$),

$$\left(\chi_{[-a,a]}\right)^\wedge(\xi) = 2a\,\widehat{f}(2a\xi) = 2a\operatorname{sinc}(2a\xi) = \frac{\sin(2\pi a\xi)}{\pi\xi}. \tag{21.7}$$

For $a = \frac{1}{2}$ we recover $\widehat{f}$, and at $\xi = 0$ we get $2a$, the integral of $\chi_{[-a,a]}$.

Translations. Let $\alpha < \beta$. The characteristic function of $[\alpha, \beta]$ is the translation by the midpoint $m = \frac{\alpha+\beta}{2}$ of $\chi_{[-a,a]}$, $a = \frac{\beta-\alpha}{2}$: $\chi_{[\alpha,\beta]} = \tau_m \chi_{[-a,a]}$. By property (ii) of Proposition 20.6 (translations become modulations),

$$\left(\chi_{[\alpha,\beta]}\right)^\wedge(\xi) = e^{-\pi i(\alpha+\beta)\xi}\,\frac{\sin(\pi(\beta-\alpha)\xi)}{\pi\xi}. \tag{21.8}$$

Cubes and rectangles. Let $Q = I_1 \times \ldots \times I_d$ be a rectangle, a product of bounded intervals,

so that $\chi_Q(x) = \prod_{j=1}^d \chi_{I_j}(x_j)$. More generally, call f a *tensor product* ("separation of variables") if

$$f(x) = f_1(x_1) \cdots f_d(x_d) \tag{21.9}$$

with $f_j \in L^1(\mathbb{R})$. Then $f \in L^1(\mathbb{R}^d)$, and since the exponential splits as a product

$$e^{-2\pi i x \cdot \xi} = \prod_j e^{-2\pi i x_j \xi_j}, \tag{21.10}$$

Fubini's theorem gives

$$\widehat{f}(\xi) = \prod_{j=1}^d \int_{\mathbb{R}} f_j(x_j)\, e^{-2\pi i x_j \xi_j}\, dx_j = \prod_{j=1}^d \widehat{f_j}(\xi_j). \tag{21.11}$$

In particular, for the cube $Q = \left[-\frac{1}{2}, \frac{1}{2}\right]^d$,

$$\widehat{\chi_Q}(\xi) = \prod_{j=1}^d \frac{\sin(\pi \xi_j)}{\pi \xi_j}, \tag{21.12}$$

and for a general rectangle $Q = \prod_j [\alpha_j, \beta_j]$,

$$\widehat{\chi_Q}(\xi) = \prod_{j=1}^d e^{-\pi i(\alpha_j + \beta_j)\xi_j} \frac{\sin(\pi(\beta_j - \alpha_j)\xi_j)}{\pi \xi_j}. \tag{21.13}$$

Lemma 21.1 (Gaussian integral)

$$\int_{\mathbb{R}} e^{-\pi x^2} dx = 1.$$

Proof. Let I be the integral; $I > 0$. By Tonelli's theorem and polar coordinates,

$$I^2 = \int_{\mathbb{R}^2} e^{-\pi(x^2+y^2)}\, dx\, dy = \int_0^{2\pi} \int_0^\infty e^{-\pi r^2} r\, dr\, d\theta = 2\pi \left[-\frac{e^{-\pi r^2}}{2\pi} \right]_0^\infty = 1. \tag{21.14}$$

■

Proposition 21.2 (Fourier transform of the Gaussian)

Let $g(x) = e^{-\pi |x|^2}$ on $\mathbb{R}^d$. Then $\widehat{g} = g$. More generally, for $\lambda > 0$ and $g_\lambda(x) = e^{-\pi \lambda |x|^2}$,

$$\widehat{g_\lambda}(\xi) = \lambda^{-\frac{d}{2}} e^{-\frac{\pi |\xi|^2}{\lambda}}. \tag{21.15}$$

So the Fourier transform of a Gaussian is still a Gaussian, and for $\lambda = 1$ the Gaussian g is a fixed point of the Fourier transform.

Proof. **Step 1: $d = 1$ and $\lambda = 1$.** Fix $\xi \in \mathbb{R}$. We complete the square: since $-\pi(x + i\xi)^2 =$

$-\pi x^2 - 2\pi i x\xi + \pi\xi^2$,

$$e^{-2\pi i x\xi} e^{-\pi x^2} = e^{-\pi\xi^2}\, e^{-\pi(x+i\xi)^2}. \tag{21.16}$$

Hence

$$\widehat{g}\,(\xi) = e^{-\pi\xi^2} \int_{\mathbb{R}} e^{-\pi(x+i\xi)^2} \mathrm{d}x, \tag{21.17}$$

and the last integral is the integral of the entire function $F\,(z) = e^{-\pi z^2}$ along the horizontal line $\operatorname{Im} z = \xi$. We compare it with the integral along the real axis by Cauchy's theorem applied to the rectangle Γ_R with vertices $-R, R, R + i\xi, -R + i\xi$ (a contour), for which $\oint_{\Gamma_R} F = 0$. On the vertical sides $z = \pm R + iy$ with y between 0 and ξ,

$$|F\,(\pm R + iy)| = e^{-\pi\left(R^2 - y^2\right)} \le e^{-\pi\left(R^2 - \xi^2\right)}, \tag{21.18}$$

so each vertical side contributes at most $|\xi|\, e^{-\pi\left(R^2-\xi^2\right)} \to 0$ as $R \to \infty$. Therefore

$$\int_{\mathbb{R}} e^{-\pi(x+i\xi)^2} \mathrm{d}x = \int_{\mathbb{R}} e^{-\pi x^2} \mathrm{d}x = 1 \tag{21.19}$$

by Lemma 21.1, and $\widehat{g}\,(\xi) = e^{-\pi\xi^2}$.

Step 2: d arbitrary, $\lambda = 1$. The Gaussian is a tensor product, $g\,(x) = \prod_j e^{-\pi x_j^2}$, so by Example 21.1 and Step 1,

$$\widehat{g}\,(\xi) = \prod_j e^{-\pi\xi_j^2} = e^{-\pi|\xi|^2}. \tag{21.20}$$

Step 3: general λ. We have $g_\lambda = g \circ T$ with $Tx = \sqrt{\lambda}\, x$, $|\det T|^{-1} = \lambda^{-\frac{d}{2}}$ and $(T^*)^{-1}\,\xi = \frac{\xi}{\sqrt{\lambda}}$. By property (iii) of Proposition 20.6,

$$\widehat{g_\lambda}\,(\xi) = \lambda^{-\frac{d}{2}} \widehat{g}\left(\frac{\xi}{\sqrt{\lambda}}\right) = \lambda^{-\frac{d}{2}} e^{-\frac{\pi|\xi|^2}{\lambda}}. \tag{21.21}$$

■

Remark 21.4 (Gaussians and approximations of the identity)**.** Let

$$G_t\,(x) = t^{-d} g\left(\frac{x}{t}\right) = t^{-d} e^{-\frac{\pi|x|^2}{t^2}} \tag{21.22}$$

for $t > 0$ (the dilations of g as in the theory of approximations of the identity, with $\Phi = g$ and $\varepsilon = t$); G_t is called the Gauss–Weierstrass (or heat) kernel, $\int_{\mathbb{R}^d} G_t = 1$, and G_t is radial and non-increasing. By (21.15) with $\lambda = t^2$,

$$\widehat{g}_{t^2} = G_t. \tag{21.23}$$

So the Fourier transforms of the Gaussians g_λ, $\lambda > 0$, form a family of approximations of the identity (as $\lambda \to 0$). Another important example, the Fourier transform of $e^{-2\pi\lambda|x|}$, is outlined in Exercise 21.1 below.

§21.iii. The Riemann–Lebesgue lemma

We denote by $C_0\left(\mathbb{R}^d\right)$ the space of continuous functions u with $u(\xi) \to 0$ as $|\xi| \to \infty$. The next result is our first important qualitative result: it says that the Fourier transform of an L^1 function is always continuous and vanishes at infinity.

Lemma 21.3 (Riemann–Lebesgue lemma)
If $f \in L^1\left(\mathbb{R}^d\right)$, then $\widehat{f} \in C_0\left(\mathbb{R}^d\right)$.

Proof. Continuity is Lemma 20.5: the integrand $f(x)\, e^{-2\pi i x\cdot\xi}$ is continuous in ξ and dominated by $|f| \in L^1$, so Dominated Convergence Theorem applies.

Let $Q = \prod_j [\alpha_j, \beta_j]$ be a rectangle. By (21.13) and $|\sin| \le 1$, the j-th factor of $\widehat{\chi}_Q$ is

$$\left| e^{-\pi i(\alpha_j+\beta_j)\xi_j} \frac{\sin\left(\pi\left(\beta_j-\alpha_j\right)\xi_j\right)}{\pi\xi_j} \right| \le \frac{1}{\pi\left|\xi_j\right|}. \tag{21.24}$$

If $|\xi|$ is large, some coordinate satisfies $|\xi_j| \ge \frac{|\xi|}{\sqrt{d}}$, so

$$\left|\widehat{\chi}_Q(\xi)\right| \le \frac{\sqrt{d}}{\pi|\xi|} \prod_{k\neq j} \left(\beta_k-\alpha_k\right) \to 0 \quad \text{as } |\xi| \to \infty. \tag{21.25}$$

By linearity of the Fourier transform, $\widehat{g} \in C_0$ for every step function g (finite linear combination of characteristic functions of rectangles). Linear combinations of functions vanishing at infinity vanish at infinity: if the property holds for g_1, g_2, it holds for $c_1 g_1 + c_2 g_2$.

Step functions are dense in $L^1\left(\mathbb{R}^d\right)$. Let $\varepsilon > 0$ and choose a step function g with

$$\|f-g\|_{L^1} < \frac{\varepsilon}{2} \tag{21.26}$$

There is $R > 0$ such that for $|\xi| > R$,

$$\left|\widehat{g}(\xi)\right| < \frac{\varepsilon}{2}. \tag{21.27}$$

Then, by linearity and property (i) of Proposition 20.6, for $|\xi| > R$,

$$\left|\widehat{f}(\xi)\right| \le \left|(f-g)^{\wedge}(\xi)\right| + \left|\widehat{g}(\xi)\right| \le \|f-g\|_{L^1} + \left|\widehat{g}(\xi)\right| < \varepsilon. \tag{21.28}$$

So $\widehat{f}(\xi) \to 0$ as $|\xi| \to \infty$. ∎

Remark 21.5 (Which functions are Fourier transforms?)**.** The Riemann–Lebesgue lemma gives a necessary condition: if $\mathcal{F}$ denotes the Fourier transform operator on L^1, then $\mathcal{F}\left(L^1\left(\mathbb{R}^d\right)\right) \subsetneq C_0\left(\mathbb{R}^d\right)$. The inclusion is strict (see Exercise 21.4). For instance, the constant function 1, which is continuous but does not vanish at infinity, is not the Fourier transform of any L^1 function. It is the Fourier transform of the Dirac delta, which is a measure and not a function: formally, $\widehat{\delta}_0(\xi) = \int e^{-2\pi i x\cdot\xi}\, \mathrm{d}\delta_0(x) = 1$, for every ξ. We will not pursue this here.

§21.iv. The multiplication formula and the inversion theorem

We now investigate the inverse problem: how to recover f from $\widehat{f}$? For this we define a closely related operator, the **inverse Fourier transform** of $f \in L^1(\mathbb{R}^d)$,

$$\check{f}(\xi) = \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} f(x)\,\mathrm{d}x = \widehat{f}(-\xi), \tag{21.29}$$

which is the same as the Fourier transform with a plus sign in the exponent. In particular $\check{f} \in C_0$ by the Riemann–Lebesgue lemma. We will see that if f and $\widehat{f}$ are both in L^1, then $f = \left(\widehat{f}\right)^\vee$ almost everywhere. First, a useful formula.

Theorem 21.4 (Multiplication formula)

If $f, g \in L^1(\mathbb{R}^d)$, then

$$\int_{\mathbb{R}^d} \widehat{f}(x)\, g(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} f(x)\,\widehat{g}(x)\,\mathrm{d}x. \tag{21.30}$$

Proof. Both integrals are well defined since $\widehat{f}, \widehat{g}$ are bounded and $f, g \in L^1$. We just open up both sides. The LHS is

$$\int_{\mathbb{R}^d} \widehat{f}(x)\, g(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(y)\, e^{-2\pi i x\cdot y}\mathrm{d}y\right) g(x)\,\mathrm{d}x \tag{21.31}$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-2\pi i x\cdot y} g(x)\, f(y)\,\mathrm{d}x\mathrm{d}y, \tag{21.32}$$

where we used Fubini's theorem (the absolute value of the integrand is $|g(x)|\,|f(y)|$, which is integrable on $\mathbb{R}^d \times \mathbb{R}^d$). Opening up the RHS,

$$\int_{\mathbb{R}^d} f(y)\,\widehat{g}(y)\,\mathrm{d}y = \int_{\mathbb{R}^d} f(y) \left(\int_{\mathbb{R}^d} g(x)\, e^{-2\pi i y\cdot x}\mathrm{d}x\right)\mathrm{d}y \tag{21.33}$$

$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-2\pi i x\cdot y} g(x)\, f(y)\,\mathrm{d}x\mathrm{d}y, \tag{21.34}$$

by Fubini's theorem again. The two sides are equal to the same double integral. ∎

Theorem 21.5 (Inversion formula)

If $f, \widehat{f} \in L^1(\mathbb{R}^d)$, then

$$f(x) = \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi}\,\widehat{f}(\xi)\,\mathrm{d}\xi = \left(\widehat{f}\right)^\vee(x) \tag{21.35}$$

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

Let us first see the idea. By the multiplication formula we can move the Fourier transform from f to the kernel $e^{2\pi i x\cdot\xi}$. The Fourier transform of this oscillatory function is, morally, a Dirac delta concentrated at x, and integrating f against it would give $f(x)$. But we cannot

apply the multiplication formula to $e^{2\pi i x\cdot\xi}$, which is not in L^1. The way out is to multiply it by a fast-decaying Gaussian, so that it becomes integrable and its Fourier transform becomes a function, namely an approximation of the identity; at the end, we let the Gaussian disappear.

Proof. Fix $x \in \mathbb{R}^d$ and $t > 0$. Define

$$\phi(\xi) = e^{2\pi i x\cdot\xi} e^{-\pi t^2|\xi|^2} \in L^1\left(\mathbb{R}^d\right). \tag{21.36}$$

By Proposition 21.2 (with $\lambda = t^2$), the Fourier transform of ϕ is

$$\widehat{\phi}(y) = \int_{\mathbb{R}^d} e^{-2\pi i(y-x)\cdot\xi} e^{-\pi t^2|\xi|^2} \mathrm{d}\xi \tag{21.37}$$

$$= \widehat{g}_{t^2}(y-x) = t^{-d} e^{-\frac{\pi|x-y|^2}{t^2}} \tag{21.38}$$

$$= G_t(x-y), \tag{21.39}$$

where G_t is the Gauss–Weierstrass kernel of Remark 21.4. Applying the Multiplication formula to the L^1 functions f and ϕ,

$$\int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} e^{-\pi t^2|\xi|^2} \widehat{f}(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}^d} \phi(\xi)\,\widehat{f}(\xi)\,\mathrm{d}\xi = \int_{\mathbb{R}^d} \widehat{\phi}(y)\, f(y)\,\mathrm{d}y$$

$$= \int_{\mathbb{R}^d} f(y)\, G_t(x-y)\,\mathrm{d}y \tag{21.40}$$

$$= f * G_t(x). \tag{21.41}$$

Now we let $t \to 0$ on both sides.

LHS. Since $\widehat{f} \in L^1$ by hypothesis, and $e^{-\pi t^2|\xi|^2} \to 1$ as $t \to 0$ for each ξ with

$$\left| e^{2\pi i x\cdot\xi} e^{-\pi t^2|\xi|^2} \widehat{f}(\xi) \right| \le \left| \widehat{f}(\xi) \right|, \tag{21.42}$$

the Dominated Convergence Theorem gives, for *every* x,

$$\int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} e^{-\pi t^2|\xi|^2} \widehat{f}(\xi)\,\mathrm{d}\xi \to \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} \widehat{f}(\xi)\,\mathrm{d}\xi. \tag{21.43}$$

(This is where the hypothesis $\widehat{f} \in L^1$ is used: we need it to apply Dominated Convergence Theorem.)

RHS. G_t is the family Φ_t associated with $\Phi = g$, the Gaussian, which is non-negative, radial non-increasing, integrable (so it is its own radial non-increasing majorant) and $\int g = 1$. By Theorem 20.1 (Approximations of the identity: pointwise convergence), $f * G_t(x) \to f(x)$ for almost every x.

Comparing the two limits in (21.41), we obtain

$$f(x) = \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} \widehat{f}(\xi)\,\mathrm{d}\xi, \qquad \text{for almost every } x \in \mathbb{R}^d. \tag{21.44}$$

■

Remark 21.6. Since $\widehat{f} \in L^1$, the function $\left(\widehat{f}\right)^\vee$ is continuous and vanishes at infinity, by Riemann–Lebesgue lemma. So if $f, \widehat{f} \in L^1$, then f has a representative (equal to f almost everywhere) which is continuous and belongs to C_0.

The proof shows how to force the convergence of the integral by multiplying the integrand by a fast decaying kernel (here the Gaussian). It also yields the following corollary, in which we do not assume that $\widehat{f} \in L^1$ but we assume that $\widehat{f}$ is non-negative and that f is continuous at 0. We need a lemma about convergence at a point of continuity.

Lemma 21.6

Let $f \in L^1\left(\mathbb{R}^d\right)$ be a function defined at every point which is continuous at the point x_0. Then $f * G_t(x_0) \to f(x_0)$ as $t \to 0$.

Proof. Let $\varepsilon > 0$ and choose $\delta > 0$ such that

$$|f(x_0 - y) - f(x_0)| < \varepsilon \tag{21.45}$$

for $|y| < \delta$. Since $\int G_t = 1$,

$$\begin{aligned} |f * G_t(x_0) - f(x_0)| &\leq \int_{|y|<\delta} |f(x_0 - y) - f(x_0)| \, G_t(y) \, \mathrm{d}y \\ &\quad + \int_{|y|\geq\delta} (|f(x_0 - y)| + |f(x_0)|) \, G_t(y) \, \mathrm{d}y \\ &\leq \varepsilon + \|f\|_{L^1} \sup_{|y|\geq\delta} G_t(y) + |f(x_0)| \int_{|y|\geq\delta} G_t(y) \, \mathrm{d}y. \end{aligned}$$

As $t \to 0$,

$$\sup_{|y|\geq\delta} G_t(y) = t^{-d} e^{-\frac{\pi\delta^2}{t^2}} \to 0 \tag{21.46}$$

and

$$\int_{|y|\geq\delta} G_t(y) \, \mathrm{d}y = \int_{|z|\geq\frac{\delta}{t}} g(z) \, \mathrm{d}z \to 0. \tag{21.47}$$

Hence,

$$\limsup_{t\to 0} |f * G_t(x_0) - f(x_0)| \leq \varepsilon, \tag{21.48}$$

for every $\varepsilon > 0$. In other words, the limsup is 0, and consequently, $f * G_t(x_0) \to f(x_0)$ as $t \to 0$. ■

Corollary 21.7

Suppose that $f \in L^1\left(\mathbb{R}^d\right)$, that $\widehat{f} \geq 0$, and that f is continuous at $x = 0$ (so 0 belongs to the Lebesgue set of f). Then $\widehat{f} \in L^1\left(\mathbb{R}^d\right)$ and therefore

$$f(x) = \int_{\mathbb{R}^d} e^{2\pi i x\cdot\xi} \widehat{f}(\xi)\, \mathrm{d}\xi \tag{21.49}$$

for almost every $x \in \mathbb{R}^d$. In particular,

$$f(0) = \int_{\mathbb{R}^d} \widehat{f}(\xi)\, \mathrm{d}\xi. \tag{21.50}$$

Proof. The identity (21.41) was proved for every $f \in L^1$ and every x (it only used the Multiplication formula), so for $x = 0$,

$$\int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{-\pi t^2|\xi|^2} \mathrm{d}\xi = f * G_t(0) \to f(0) \quad \text{as } t \to 0 \tag{21.51}$$

by Lemma 21.6. Since $\widehat{f} \geq 0$ and $e^{-\pi t^2|\xi|^2}$ increases to 1 as $t \to 0$, Monotone Convergence Theorem says that

$$\lim_{t\to 0} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{-\pi t^2|\xi|^2} \mathrm{d}\xi = \int_{\mathbb{R}^d} \widehat{f}(\xi) \left(\lim_{t\to 0} e^{-\pi t^2|\xi|^2}\right) \mathrm{d}\xi = \int_{\mathbb{R}^d} \widehat{f}(\xi)\, \mathrm{d}\xi. \tag{21.52}$$

Alternatively, one can use Fatou's Lemma, which gives

$$\int_{\mathbb{R}^d} \widehat{f}(\xi)\, \mathrm{d}\xi \leq \liminf_{t\to 0} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{-\pi t^2|\xi|^2} \mathrm{d}\xi = \lim_{t\to 0} f * G_t(0) = f(0). \tag{21.53}$$

Therefore,

$$\int_{\mathbb{R}^d} \widehat{f}(\xi)\, \mathrm{d}\xi = \lim_{t\to 0} \int_{\mathbb{R}^d} \widehat{f}(\xi)\, e^{-\pi t^2|\xi|^2} \mathrm{d}\xi = \lim_{t\to 0} f * G_t(0) = f(0) < \infty; \tag{21.54}$$

so $\widehat{f} \in L^1$ and the equality $f(0) = \int \widehat{f}$ holds. Finally, Inversion formula applies, since $f, \widehat{f} \in L^1$. ∎

§21.v. The L^2 theory

There is a simple and elegant way to extend the theory to the L^2-context. The key ingredient is the following result, for $f \in L^1\left(\mathbb{R}^d\right) \cap L^2\left(\mathbb{R}^d\right)$.

Theorem 21.8 (Isometry property)

If $f \in L^1\left(\mathbb{R}^d\right) \cap L^2\left(\mathbb{R}^d\right)$, then $\left\|\widehat{f}\right\|_{L^2} = \|f\|_{L^2}$.

Proof. Let $g(x) = \overline{f(-x)}$ and $h = f * g$. Since $f, g \in L^1$, we know from Young's inequality (with $p = q = r = 1$) that $h \in L^1$, and so $\widehat{h} = \widehat{f}\widehat{g}$ by property (iv) of Proposition 20.6. Now

$$\widehat{g}(\xi) = \int_{\mathbb{R}^d} \overline{f(-x)} e^{-2\pi i x\cdot\xi}\, \mathrm{d}x = \int_{\mathbb{R}^d} \overline{f(y)} e^{2\pi i y\cdot\xi}\, \mathrm{d}y = \overline{\widehat{f}(\xi)}, \tag{21.55}$$

so $\widehat{h} = \left|\widehat{f}\right|^2 \geq 0$. Moreover, since $f, g \in L^2$, Proposition 19.4 (with $p = q = 2$) shows that h is bounded and uniformly continuous; in particular h is continuous at 0. By Corollary 21.7, $\widehat{h} \in L^1$ and

$$h(0) = \int_{\mathbb{R}^d} \widehat{h}(\xi)\, \mathrm{d}\xi. \tag{21.56}$$

But

$$h(0) = \int_{\mathbb{R}^d} f(y)\, g(-y)\, \mathrm{d}y = \int_{\mathbb{R}^d} f(y)\, \overline{f(y)} \mathrm{d}y = \int_{\mathbb{R}^d} |f(x)|^2\, \mathrm{d}x, \tag{21.57}$$

and therefore

$$\int_{\mathbb{R}^d} |f(x)|^2\, \mathrm{d}x = h(0) = \int_{\mathbb{R}^d} \widehat{h}(\xi)\, \mathrm{d}\xi = \int_{\mathbb{R}^d} \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi. \tag{21.58}$$

■

Now, we can define the Fourier transform on $L^2(\mathbb{R}^d)$. The space $L^1 \cap L^2$ is dense in L^2, because, for example, $L^1 \cap L^2$ contains all compactly supported continuous functions. By Theorem 21.8 the Fourier transform is an isometry for the L^2-norm on $L^1 \cap L^2$.

Definition 21.1 (Fourier transform on $L^2(\mathbb{R}^d)$)**.** We define $\mathcal{F} : L^2 \to L^2$ as follows: Let $f \in L^2$ and let $g_n \in L^1 \cap L^2$ with $g_n \to f$ in L^2. Then $\{g_n\}$ is Cauchy in L^2, and so is $\{\widehat{g}_n\}$, because $\|\widehat{g}_n - \widehat{g}_m\|_{L^2} = \|g_n - g_m\|_{L^2}$. Hence $\widehat{g}_n$ converges in L^2 to a limit, say $\widehat{f}$, which we *define* to be

$$\mathcal{F}f := \widehat{f} = \lim_{n\to\infty} \widehat{g}_n. \tag{21.59}$$

For the approximating sequence, we can, for example, take $g_n = f\chi_{B(0,n)}$, which is in L^1 by Cauchy–Schwarz inequality:

$$\int_{\mathbb{R}^d} |f\chi_{B(0,n)}| \leq \|f\|_2 \left\|\chi_{B(0,n)}\right\|_2 < \infty. \tag{21.60}$$

Furthermore, the definition is well-defined, i.e. the limit $g_n \to \widehat{f}$ does not depend on the approximating sequence. To see that, if h_n is another such sequence converging to f,

$$\left\|\widehat{g}_n - \widehat{h}_n\right\|_{L^2} = \left\|\widehat{g_n - h_n}\right\|_{L^2} = \|g_n - h_n\|_{L^2} \to 0. \tag{21.61}$$

So

$$\lim_{n\to\infty} \widehat{g}_n = \lim_{n\to\infty} \widehat{h}_n, \tag{21.62}$$

so the limit doesn't depend on the approximating sequence.

Also, if $f \in L^1 \cap L^2$, then $g_n = f\chi_{B(0,n)} \to f$ both in L^1 and in L^2, so $\widehat{g}_n \to \widehat{f}$ uniformly and in L^2, and the two definitions agree. So $\mathcal{F} : L^2 \to L^2$ is the unique bounded extension of the

Fourier transform from the dense subspace $L^1 \cap L^2$, it is linear, and $\|\mathcal{F}f\|_{L^2} = \|f\|_{L^2}$ for $f \in L^2$. In particular, if χ_k is the characteristic function of the open ball of radius $k \in \mathbb{Z}_{>0}$ and $f \in L^2$, then $f_k = f\chi_k$ satisfy $\widehat{f} = \lim_{k\to\infty} \widehat{f_k}$ in L^2.

Remark 21.7. The natural domain of definition of the Fourier transform is L^2. The formula $\int f e^{-2\pi i x\cdot\xi}\,\mathrm{d}x$ does not make sense for general $f \in L^2$, but the isometry property tells us how to define $\widehat{f}$ as a limit.

Before the main theorem, observe that Multiplication formula remains valid for $f, g \in L^2$.

Proposition 21.9 (Multiplication formula for L^2)

Let $f, g \in L^2\left(\mathbb{R}^d\right)$, then

$$\int_{\mathbb{R}^d} \widehat{f}\,(x)\, g\,(x)\ \mathrm{d}x = \int_{\mathbb{R}^d} f\,(x)\, \widehat{g}\,(x)\ \mathrm{d}x. \tag{21.63}$$

Proof. Let f_n, g_n be sequences in $L^1 \cap L^2$ such that $f_n \to f$ and $g_n \to g$ in L^2. Then by Multiplication formula for L^1, we have

$$\int_{\mathbb{R}^d} \widehat{f_n}\,(x)\, g_n\,(x)\ \mathrm{d}x = \int_{\mathbb{R}^d} f_n\,(x)\, \widehat{g_n}\,(x)\ \mathrm{d}x. \tag{21.64}$$

Now, by Cauchy–Schwarz inequality (or Hölder's inequality) and Isometry property,

$$\begin{aligned}
\left|\int_{\mathbb{R}^d} \widehat{f_n} g_n - \int_{\mathbb{R}^d} \widehat{f} g\right| &\le \int_{\mathbb{R}^d} \left|\widehat{f_n}\right| |g_n - g| + \int_{\mathbb{R}^d} \left|\widehat{f_n} - \widehat{f}\right| |g| && (21.65)\\
&\le \|f_n\|_2 \|g_n - g\|_2 + \left\|\widehat{f_n} - \widehat{f}\right\|_2 \|g\|_2 && (21.66)\\
&= \|f\|_2 \|g_n - g\|_2 + \left\|\widehat{f_n} - \widehat{f}\right\|_2 \|g\|_2 \to 0 && (21.67)
\end{aligned}$$

as $n \to \infty$. Similarly,

$$\left|\int_{\mathbb{R}^d} f_n \widehat{g_n} - \int_{\mathbb{R}^d} f\widehat{g}\right| \to 0 \text{ as } n \to \infty. \tag{21.68}$$

Now, taking $n \to \infty$ in (21.64), we are done! ∎

Proposition 21.10

If $f \in L^2$, and define $\widetilde{f}$ to be the function $\widetilde{f}\,(x) = f\,(-x)$. Then

$$\mathcal{F}\left(\widetilde{f}\right) = \overline{\mathcal{F}\left(\overline{f}\right)}. \tag{21.69}$$

Proof. Let f_n be a sequence in $L^1 \cap L^2$ such that $f_n \to f$ in L^2. Then $\widetilde{f_n}$ is a sequence in $L^1 \cap L^2$ such that $\widetilde{f_n} \to \widetilde{f}$ in L^2. Then by (21.55),

$$\widehat{\widehat{f_n}} = \widetilde{\widehat{f_n}}. \tag{21.70}$$

Taking $n \to \infty$, we have that

$$\mathcal{F}\left(\widetilde{f}\right) = \lim_{n\to\infty} \widehat{\widetilde{f_n}} = \lim_{n\to\infty} \widetilde{\widehat{f_n}} = \overline{\lim_{n\to\infty} \widehat{\widetilde{f_n}}} = \overline{\mathcal{F}\left(\overline{f}\right)}. \tag{21.71}$$

■

Theorem 21.11 (Plancherel's theorem)

The operator $\mathcal{F} : L^2\left(\mathbb{R}^d\right) \to L^2\left(\mathbb{R}^d\right)$ is unitary, i.e. linear, isometric and surjective. Moreover, its inverse $\mathcal{F}^{-1}$ is given by

$$\left(\mathcal{F}^{-1} g\right)(x) = \mathcal{F} g\left(-x\right), \qquad g \in L^2\left(\mathbb{R}^d\right). \tag{21.72}$$

Proof. ***Linear and isometric.*** This has been established above (construction of $\mathcal{F}$ and Isometry property). In particular, $\mathcal{F}$ preserves the inner product $(u, v) = \int u\bar{v}$ (by Parallelogram law), and its image is a closed subspace of L^2 (the isometric image of a complete space is complete).

Surjective. Suppose that $\mathcal{F}\left(L^2\right) \subsetneq L^2$. Since the image is a closed subspace of the Hilbert space L^2, there exists $g \in L^2$, $g \neq 0$, orthogonal to the image, i.e.

$$\int_{\mathbb{R}^d} \widehat{f}\,\bar{g} = 0. \tag{21.73}$$

for all $f \in L^2$. By Multiplication formula for L^2 (applied to f and $\bar{g} \in L^2$),

$$0 = \int_{\mathbb{R}^d} \widehat{f}\,\bar{g} = \int_{\mathbb{R}^d} f\,\widehat{\bar{g}}, \quad \text{for all } f \in L^2. \tag{21.74}$$

Choosing $f = \widetilde{\bar{\widehat{g}}} \in L^2$, we get

$$\int_{\mathbb{R}^d} \left|\widehat{\bar{g}}\right|^2 = 0 \tag{21.75}$$

so $\widehat{\bar{g}} = 0$, and then

$$\|g\|_{L^2} = \|\bar{g}\|_{L^2} = \left\|\widehat{\bar{g}}\right\|_{L^2} = 0 \tag{21.76}$$

i.e. $g = 0$, a contradiction.

The inverse. Let $\widetilde{g}\left(x\right) = g\left(-x\right)$. As noted above in Proposition 21.10, $\mathcal{F}\widetilde{g} = \overline{\mathcal{F}\left(\bar{g}\right)}$.Now, for $f, g \in L^2$, using Multiplication formula for L^2 and the fact that $\mathcal{F}$ is an isometry (hence preserves the inner product), we have

$$\left(f, \mathcal{F}\left(\widetilde{\mathcal{F}g}\right)\right) = \int_{\mathbb{R}^d} f\,\overline{\mathcal{F}\left(\widetilde{\mathcal{F}g}\right)} = \int_{\mathbb{R}^d} f\,\mathcal{F}\left(\overline{\mathcal{F}g}\right) \tag{21.77}$$

$$= \int_{\mathbb{R}^d} \mathcal{F}f\,\overline{\mathcal{F}g} = \left(\mathcal{F}f, \mathcal{F}g\right) \tag{21.78}$$

$$= \left(f, g\right). \tag{21.79}$$

Here we used $\overline{\mathcal{F}(\widetilde{u})} = \overline{\overline{\mathcal{F}(\bar{u})}} = \mathcal{F}(\bar{u})$ with $u = \mathcal{F}g$. Since (21.79) holds for every $f \in L^2$, we get

$$\mathcal{F}\left(\widetilde{\mathcal{F}g}\right) = g, \tag{21.80}$$

for every g, i.e.

$$\mathcal{F} \circ \mathcal{R} \circ \mathcal{F} = I \tag{21.81}$$

where $\mathcal{R}u(x) = u(-x)$. Thus $\mathcal{R}\mathcal{F}$ is a right inverse of the bijection $\mathcal{F}$, hence it is its inverse:

$$\mathcal{F}^{-1}g = \mathcal{R}\mathcal{F}g. \tag{21.82}$$

In other words,

$$\left(\mathcal{F}^{-1}g\right)(x) = \mathcal{F}g(-x). \tag{21.83}$$

■

Remark 21.8. When $\widehat{g} \in L^1$ as well, this formula is the inversion formula of Theorem 21.5: $\mathcal{F}^{-1}$ is given by the inverse Fourier transform $(-)^\vee$.

§21.vi. The Fourier transform in L^p for $1 \le p \le 2$

Once the Fourier transform is defined in L^1 and L^2, we can extend its definition to the larger class of functions

$$L^1\left(\mathbb{R}^d\right) + L^2\left(\mathbb{R}^d\right) = \left\{f_1 + f_2 \ : \ f_1 \in L^1,\ f_2 \in L^2\right\}. \tag{21.84}$$

If $f = f_1 + f_2$ as above, we define $\widehat{f} = \widehat{f_1} + \widehat{f_2}$. This definition does not depend on the decomposition. Indeed, if

$$f = f_1 + f_2 = g_1 + g_2 \tag{21.85}$$

with $f_i, g_i \in L^i$, $i = 1, 2$, then

$$h = f_1 - g_1 = g_2 - f_2 \in L^1 \cap L^2 \tag{21.86}$$

and on $L^1 \cap L^2$ the L^1-transform and the L^2-transform agree, so

$$\widehat{h} = \widehat{f_1} - \widehat{g_1} = \widehat{g_2} - \widehat{f_2}, \tag{21.87}$$

which implies

$$\widehat{f_1} + \widehat{f_2} = \widehat{g_1} + \widehat{g_2}. \tag{21.88}$$

For $1 \le p \le 2$ we have $L^p \subseteq L^1 + L^2$: given $f \in L^p$, write

$$f = f\chi_{\{|f|>1\}} + f\chi_{\{|f|\le 1\}}. \tag{21.89}$$

Notice that the first term is in L^1 and the second one is in L^2. Because for the first term, $|f| > 1$, so for $p \geq 1$, we have

$$|f| \leq |f|^p \,, \tag{21.90}$$

so as a result,

$$\left\| f\chi_{\{|f|>1\}} \right\|_1 = \int_{\{|f|>1\}} |f| \leq \int_{\{|f|>1\}} |f|^p \leq \|f\|_p^p \,. \tag{21.91}$$

For the second term, since $|f| \leq 1$ and $p \leq 2$, we have

$$|f|^2 \leq |f|^p \,, \tag{21.92}$$

so as a result,

$$\left\| f\chi_{\{|f|\leq 1\}} \right\|_2^2 = \int_{\{|f|\leq 1\}} |f|^2 \leq \int_{\{|f|\leq\}} |f|^p \leq \|f\|_p^p \,. \tag{21.93}$$

So $L^p \subseteq L^1 + L^2$, and we can define the Fourier transform on the L^p-spaces, $1 \leq p \leq 2$. The following result is another classical inequality in Fourier analysis.

Theorem 21.12 (Hausdorff–Young inequality)

If $f \in L^p\left(\mathbb{R}^d\right)$ with $1 \leq p \leq 2$ and $\frac{1}{p} + \frac{1}{p'} = 1$, then

$$\left\| \widehat{f} \right\|_{L^{p'}} \leq \|f\|_{L^p} \,. \tag{21.94}$$

The sharp form of this inequality is due to Beckner (see Exercise 21.9). The result follows from property (i) of Proposition 20.6, $\left\| \widehat{f} \right\|_{L^\infty} \leq \|f\|_{L^1}$, together with the isometry property, $\left\| \widehat{f} \right\|_{L^2} = \|f\|_{L^2}$, by the Riesz–Thorin interpolation theorem. For completeness, we write out the interpolation argument (the proof of Riesz–Thorin in this special case), which uses Hadamard's three lines lemma from complex analysis.

Lemma 21.13 (Hadamard's three lines lemma)

Let Φ be a bounded continuous function on the closed strip $\{0 \leq \operatorname{Re} z \leq 1\}$, holomorphic in the open strip, with $|\Phi(it)| \leq M_0$ and $|\Phi(1+it)| \leq M_1$ for all $t \in \mathbb{R}$. Then

$$|\Phi(\theta + it)| \leq M_0^{1-\theta} M_1^\theta \tag{21.95}$$

for all $0 \leq \theta \leq 1$ and $t \in \mathbb{R}$.

(This is a standard consequence of the maximum modulus principle.)

Proof of Hausdorff–Young inequality. For $p = 1$ the statement is property (i) of Proposition 20.6, and for $p = 2$ it is the isometry property. Let $1 < p < 2$, so $2 < p' < \infty$, and define $\theta \in (0,1)$ by $\frac{1}{p} = 1 - \frac{\theta}{2}$ (the interpolation between the endpoints $(p_0, q_0) = (1, \infty)$ and $(p_1, q_1) = (2, 2)$), so that $p\left(1 - \frac{\theta}{2}\right) = 1$.

Step 1: simple functions. Let f be a simple function with $\|f\|_{L^p} = 1$, i.e. $f = \sum_{j=1}^{m} a_j \chi_{A_j}$ with $a_j = |a_j| e^{i\alpha_j}$, A_j disjoint of finite measure. Since $p' < \infty$, by Duality of the norm (and Remark 11.1),

$$\left\|\widehat{f}\right\|_{L^{p'}} = \sup\left\{\left|\int \widehat{f} g\right| : g \text{ simple},\ \|g\|_{L^p} = 1\right\}. \tag{21.96}$$

($\widehat{f}$ is bounded and in L^2, hence in $L^{p'}$.) Fix such a $g = \sum_{k=1}^{n} b_k \chi_{B_k}$, $b_k = |b_k| e^{i\beta_k}$. For z in the strip $0 \le \operatorname{Re} z \le 1$ put $\gamma(z) = p\left(1 - \frac{z}{2}\right)$ and

$$f_z = \sum_{j=1}^{m} |a_j|^{\gamma(z)} e^{i\alpha_j} \chi_{A_j}, \qquad g_z = \sum_{k=1}^{n} |b_k|^{\gamma(z)} e^{i\beta_k} \chi_{B_k}. \tag{21.97}$$

Then $f_\theta = f$ and $g_\theta = g$ since $\gamma(\theta) = 1$. Define

$$\Phi(z) = \int_{\mathbb{R}^d} (f_z)^\wedge(\xi)\, g_z(\xi)\, \mathrm{d}\xi = \sum_{j,k} |a_j|^{\gamma(z)} |b_k|^{\gamma(z)} e^{i(\alpha_j+\beta_k)} \int_{\mathbb{R}^d} \widehat{\chi}_{A_j} \chi_{B_k} \mathrm{d}\xi, \tag{21.98}$$

where each integral is finite, since $\widehat{\chi}_{A_j}$ is bounded and B_k has finite measure. So Φ is a finite sum of terms $c e^{\lambda z}$ with $\lambda \in \mathbb{R}$: it is entire and bounded in the strip.

On the line $\operatorname{Re} z = 0$: $|f_{it}| = |f|^p$, $|g_{it}| = |g|^p$, because $\operatorname{Re}\gamma(it) = p$. Thus $\|f_{it}\|_{L^1} = \|f\|_{L^p}^p = 1$ and $\|g_{it}\|_{L^1} = 1$, and by property (i) of Proposition 20.6,

$$|\Phi(it)| \le \left\|(f_{it})^\wedge\right\|_{L^\infty} \|g_{it}\|_{L^1} \le \|f_{it}\|_{L^1} \|g_{it}\|_{L^1} = 1. \tag{21.99}$$

On the line $\operatorname{Re} z = 1$: $\operatorname{Re}\gamma(1+it) = \frac{p}{2}$, so $|f_{1+it}| = |f|^{\frac{p}{2}}$, $|g_{1+it}| = |g|^{\frac{p}{2}}$, and $\|f_{1+it}\|_{L^2} = \|g_{1+it}\|_{L^2} = 1$. By the Cauchy–Schwarz inequality (or Hölder's inequality) and Plancherel's theorem,

$$|\Phi(1+it)| \le \left\|(f_{1+it})^\wedge\right\|_{L^2} \|g_{1+it}\|_{L^2} = \|f_{1+it}\|_{L^2} \|g_{1+it}\|_{L^2} = 1. \tag{21.100}$$

By the Hadamard's three lines lemma,

$$|\Phi(\theta)| \le 1, \tag{21.101}$$

i.e. $\left|\int \widehat{f} g\right| \le 1$. Taking the supremum over g,

$$\left\|\widehat{f}\right\|_{L^{p'}} \le 1 = \|f\|_{L^p}. \tag{21.102}$$

By homogeneity, $\left\|\widehat{f}\right\|_{L^{p'}} \le \|f\|_{L^p}$ for every simple function f.

Step 2: general $f \in L^p$. Simple functions (with finite-measure support) are dense in L^p, $p < \infty$. Let $f_n \to f$ in L^p be simple. Splitting $f - f_n = u_n + v_n$ as in (21.89), with $u_n \in L^1$, $v_n \in L^2$,

$$\|\widehat{u}_n\|_{L^\infty} \le \|f - f_n\|_{L^p}^p \to 0, \qquad \|\widehat{v}_n\|_{L^2} \le \|f - f_n\|_{L^p}^{\frac{p}{2}} \to 0. \tag{21.103}$$

Since $\widehat{f} - \widehat{f}_n = \widehat{u}_n + \widehat{v}_n$ (the definition of the transform on $L^1 + L^2$), we get $\widehat{f}_n \to \widehat{f}$ uniformly plus in L^2, hence a subsequence $\widehat{f}_{n_k} \to \widehat{f}$ almost everywhere (a subsequence of an L^2-convergent

sequence converges a.e.). By Fatou's Lemma and Step 1,

$$\left\|\widehat{f}\right\|_{L^{p'}} \leq \liminf_{k\to\infty} \left\|\widehat{f}_{n_k}\right\|_{L^{p'}} \leq \lim_{k\to\infty} \|f_{n_k}\|_{L^p} = \|f\|_{L^p}. \tag{21.104}$$

■

§21.vii. The Schwartz class

Consider the class of functions $f \in C^\infty\left(\mathbb{R}^d\right)$ such that

$$\sup_{x\in\mathbb{R}^d} \left|x^\alpha \left(\partial^\beta f\right)(x)\right| < \infty \tag{21.105}$$

for any multi-indices $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $\beta = (\beta_1, \ldots, \beta_d)$. In other words, this is the class of infinitely differentiable functions such that any derivative decays faster than any polynomial. This set is known as the *Schwartz class*, denoted by $\mathcal{S}\left(\mathbb{R}^d\right)$, in honor of the French mathematician Laurent Schwartz (1915–2002). A classical example of a function in the Schwartz class is the Gaussian $f(x) = e^{-\pi|x|^2}$. Moreover, $C_c^\infty\left(\mathbb{R}^d\right) \subseteq \mathcal{S}\left(\mathbb{R}^d\right)$.

Lemma 21.14 (Basic facts about $\mathcal{S}$) (a) $f \in \mathcal{S}$ if and only if $f \in C^\infty$ and, for every $N \geq 0$ and every multi-index β, $\sup_x (1+|x|)^N \left|\partial^\beta f(x)\right| < \infty$.

(b) $\mathcal{S}$ is a vector space, and if $f \in \mathcal{S}$, then $x^\alpha f \in \mathcal{S}$ and $\partial^\beta f \in \mathcal{S}$ for all multi-indices α, β.

(c) $\mathcal{S} \subseteq L^p\left(\mathbb{R}^d\right)$ for every $1 \leq p \leq \infty$, and $\mathcal{S} \subseteq C_0$. For $1 \leq p < \infty$ the Schwartz class is dense in L^p (because C_c^∞ is).

Proof. (a) Expanding $(1+|x_1|+\cdots+|x_d|)^N$ into a finite sum with positive coefficients, we see that $(1+|x|)^N \leq C_N \sum_{|\alpha|\leq N} |x^\alpha|$ (since $|x| \leq \sum_j |x_j|$); conversely $|x^\alpha| \leq |x|^{|\alpha|} \leq (1+|x|)^{|\alpha|}$.

(b) By Leibniz's rule, $\partial^\gamma\left(x^\alpha f\right)$ is a finite linear combination of terms $x^{\alpha'}\partial^{\gamma'} f$, so $x^\delta \partial^\gamma\left(x^\alpha f\right)$ is a finite combination of functions of the form $x^{\delta+\alpha'}\partial^{\gamma'} f$, which are bounded. For derivatives, $x^\alpha \partial^\gamma\left(\partial^\beta f\right) = x^\alpha \partial^{\gamma+\beta} f$ is bounded.

(c) By (a) with $N = d+1$, $|f(x)| \leq C(1+|x|)^{-d-1}$, which is in L^1 (and L^p for every p), bounded, and tends to 0 as $|x| \to \infty$. ■

The Fourier transform behaves particularly well with respect to functions in the Schwartz class, as the next proposition shows (this is Problem 77 in the notes).

Proposition 21.15 (i) If $f \in \mathcal{S}\left(\mathbb{R}^d\right)$, then $\widehat{f} \in \mathcal{S}\left(\mathbb{R}^d\right)$.

(ii) If $f, g \in \mathcal{S}\left(\mathbb{R}^d\right)$, then $f * g \in \mathcal{S}\left(\mathbb{R}^d\right)$.

Proof. (i) Let $f \in \mathcal{S}$. By Lemma 21.14 (c), $x^\alpha f \in \mathcal{S} \subseteq L^1$ for every α, so by property (v) of Proposition 20.6, $\widehat{f} \in C^k$ for every k, i.e. $\widehat{f} \in C^\infty$, and

$$\partial^\alpha \widehat{f} = \left((-2\pi i x)^\alpha f\right)^\wedge. \tag{21.106}$$

Now let $u = (-2\pi i x)^\alpha f \in \mathcal{S}$. All its derivatives $\partial^\beta u$ are in $\mathcal{S} \subseteq L^1 \cap C_0$, so property (vi) of Proposition 20.6 applies to u (with any k) and gives $\left(\partial^\beta u\right)^\wedge = (2\pi i\xi)^\beta \widehat{u}$. Hence, using property (i),

$$\left|\xi^\beta \partial^\alpha \widehat{f}(\xi)\right| = \left|\xi^\beta \widehat{u}(\xi)\right| = \frac{1}{(2\pi)^{|\beta|}} \left|\left(\partial^\beta u\right)^\wedge(\xi)\right| \le \frac{1}{(2\pi)^{|\beta|}} \left\|\partial^\beta u\right\|_{L^1} < \infty \tag{21.107}$$

uniformly in ξ. So $\sup_\xi \left|\xi^\beta \partial^\alpha \widehat{f}(\xi)\right| < \infty$ for all α, β, i.e. $\widehat{f} \in \mathcal{S}$.

(ii) Since $\partial^\beta f$ is bounded and $g \in L^1$, differentiation under the integral sign (Dominated Convergence Theorem with the mean value theorem, by Claim 20.2) shows that $f * g \in C^\infty$ with $\partial^\beta (f * g) = \left(\partial^\beta f\right) * g$. For the decay, we use the inequality $1 + |x| \le (1 + |x - y|)(1 + |y|)$. For every N,

$$\begin{aligned}(1+|x|)^N \left|\left(\partial^\beta f * g\right)(x)\right| &\le \int_{\mathbb{R}^d} (1+|x-y|)^N \left|\partial^\beta f(x-y)\right| (1+|y|)^N |g(y)|\, \mathrm{d}y \\ &\le \sup_u (1+|u|)^N \left|\partial^\beta f(u)\right| \int_{\mathbb{R}^d} (1+|y|)^N |g(y)|\, \mathrm{d}y < \infty,\end{aligned}$$

where both factors are finite by Lemma 21.14 (a) and (since $(1+|y|)^N |g(y)| \le C(1+|y|)^{-d-1}$) the integral converges. By Lemma 21.14 (a), $f * g \in \mathcal{S}$. ■

Remark 21.9 (The Fourier transform on $\mathcal{S}$)**.** By the Inversion formula, if $f \in \mathcal{S}$, then $f, \widehat{f} \in L^1$, so $f = \left(\widehat{f}\right)^\vee$ almost everywhere, and, both sides being continuous, everywhere. Together with Proposition 21.15 (i) (applied to $\check{h}(\xi) = \widehat{h}(-\xi)$, which is in $\mathcal{S}$ if h is), this shows that the Fourier transform is a bijection from $\mathcal{S}$ onto $\mathcal{S}$ with inverse $h \mapsto \check{h}$.

§21.viii. Exercises

Exercise 21.1. (i) Show that for $\beta > 0$ we have

$$e^{-\beta} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}}\, e^{-\frac{\beta^2}{4u}}\, \mathrm{d}u. \tag{21.108}$$

(ii) Prove that the Fourier transform of $f_\lambda(x) = e^{-2\pi\lambda|x|}$ in $\mathbb{R}^d$, where $\lambda > 0$, is given by

$$\widehat{f_\lambda}(\xi) = C_d \frac{\lambda}{\left(\lambda^2 + |\xi|^2\right)^{\frac{d+1}{2}}}, \tag{21.109}$$

where $C_d = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\pi^{\frac{d+1}{2}}}$.

Solution. (i) In the integral, we do the change of variables $u = x^2$, so

$$I := \int_0^\infty \frac{e^{-u}}{\sqrt{u}}\, e^{-\frac{\beta^2}{4u}}\, \mathrm{d}u = \int_0^\infty \frac{e^{-x^2}}{x}\, e^{-\frac{\beta^2}{4x^2}}\, 2x\, \mathrm{d}x = 2\int_0^\infty e^{-x^2 - \frac{\beta^2}{4x^2}}\, \mathrm{d}x \tag{21.110}$$

After completing the squares, we have

$$-x^2 - \frac{\beta^2}{4x^2} = -\left(x - \frac{\beta}{2x}\right)^2 - 2x \cdot \frac{\beta}{2x} = -\left(x - \frac{\beta}{2x}\right)^2 - \beta. \tag{21.111}$$

So we have

$$I = 2e^{-\beta} \int_0^\infty e^{-\left(x - \frac{\beta}{2x}\right)^2} \, \mathrm{d}x. \tag{21.112}$$

Claim 21.1: If $f : \mathbb{R} \to \mathbb{R}$ is an even integrable function, then for $a > 0$, we have

$$\int_0^\infty f\left(x - \frac{a}{x}\right) \, \mathrm{d}x = \int_0^\infty f(x) \, \mathrm{d}x. \tag{21.113}$$

Proof. Call the LHS integral J. After doing a change of variables $y = \frac{a}{x}$, we have

$$J = \int_\infty^0 f\left(\frac{a}{y} - y\right) \frac{-a}{y^2} \, \mathrm{d}y = \int_0^\infty \frac{a}{y^2} f\left(\frac{a}{y} - y\right) \, \mathrm{d}y. \tag{21.114}$$

Therefore, by taking average,

$$J = \frac{1}{2} \int_0^\infty \left(1 + \frac{a}{x^2}\right) f\left(x - \frac{a}{x}\right) \, \mathrm{d}x. \tag{21.115}$$

Now, let $u = x - \frac{a}{x}$, so that $\mathrm{d}u = \left(1 + \frac{a}{x^2}\right) \mathrm{d}x$, so that

$$J = \frac{1}{2} \int_{-\infty}^\infty f(u) \, \mathrm{d}u = \int_0^\infty f(x) \, \mathrm{d}x, \tag{21.116}$$

since f is even. □

Therefore,

$$\int_0^\infty e^{-\left(x - \frac{\beta}{2x}\right)^2} \, \mathrm{d}x = \int_0^\infty e^{-x^2} \, \mathrm{d}x = \frac{\sqrt{\pi}}{2}. \tag{21.117}$$

So we have

$$\frac{1}{\sqrt{\pi}} I = \frac{1}{\sqrt{\pi}} 2e^{-\beta} \frac{\sqrt{\pi}}{2} = e^{-\beta}. \tag{21.118}$$

(ii) First, notice that $e^{-2\pi\lambda|x|}$ is in L^1. Indeed, in polar coordinates,

$$\|f_\lambda\|_1 = \int_{\mathbb{R}^d} e^{-2\pi\lambda|x|} \, \mathrm{d}x = \omega_{d-1} \int_0^\infty e^{-2\pi\lambda r} \, r^{d-1} \, \mathrm{d}r \tag{21.119}$$

$$= \omega_{d-1} \int_0^\infty e^{-s} \left(\frac{s}{2\pi\lambda}\right)^{d-1} \frac{1}{2\pi\lambda} \, \mathrm{d}s, \tag{21.120}$$

after the change of variable $s = 2\pi\lambda r$. After some simplification, it becomes

$$\|f_\lambda\|_1 = \frac{\omega_{d-1}}{(2\pi\lambda)^d} \int_0^\infty e^{-s} s^{d-1} \, \mathrm{d}s = \frac{\omega_{d-1}}{(2\pi\lambda)^d} (d-1)\,! < \infty. \tag{21.121}$$

Now, using (i), we have

$$e^{-2\pi\lambda|x|} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\frac{(2\pi\lambda|x|)^2}{4u}} \mathrm{d}u. = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\pi\left(\frac{\pi\lambda^2}{u}\right)|x|^2} \mathrm{d}u. \tag{21.122}$$

So the Fourier transform is

$$\widehat{f_\lambda}(\xi) = \int_{\mathbb{R}^d} e^{-2\pi\lambda|x|} e^{-2\pi i x\cdot\xi} \,\mathrm{d}x = \frac{1}{\sqrt{\pi}} \int_{\mathbb{R}^d} \left(\int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\pi\left(\frac{\pi\lambda^2}{u}\right)|x|^2} \mathrm{d}u \right) e^{-2\pi i x\cdot\xi} \,\mathrm{d}x. \tag{21.123}$$

The integrand is bounded (in absolute value) by $\frac{e^{-u}}{\sqrt{u}} e^{-\pi\left(\frac{\pi\lambda^2}{u}\right)|x|^2}$, which is integrable on $(0,\infty)\times\mathbb{R}^d$ as one can check using Tonelli's theorem. So we can apply Fubini's theorem on (21.123) to swap the integrals.

$$\widehat{f_\lambda}(\xi) = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} \left[\int_{\mathbb{R}^d} e^{-\pi\left(\frac{\pi\lambda^2}{u}\right)|x|^2} e^{-2\pi i x\cdot\xi} \mathrm{d}x \right] \mathrm{d}u \tag{21.124}$$

Using Fourier transform of the Gaussian, we have

$$\int_{\mathbb{R}^d} e^{-\pi\left(\frac{\pi\lambda^2}{u}\right)|x|^2} e^{-2\pi i x\cdot\xi} \mathrm{d}x = \left(\frac{u}{\pi\lambda^2}\right)^{\frac{d}{2}} e^{-\frac{u|\xi|^2}{\lambda^2}}. \tag{21.125}$$

So we have

$$\widehat{f_\lambda}(\xi) = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} \left(\frac{u}{\pi\lambda^2}\right)^{\frac{d}{2}} e^{-\frac{u|\xi|^2}{\lambda^2}} \,\mathrm{d}u \tag{21.126}$$

$$= \frac{1}{\pi^{\frac{d+1}{2}}\lambda^2} \int_0^\infty u^{\frac{d-1}{2}} e^{-u\left(1+\frac{|\xi|^2}{\lambda^d}\right)} \,\mathrm{d}u. \tag{21.127}$$

Let $\alpha = 1 + \frac{|\xi|^2}{\lambda^2}$. Taking $s = u\alpha$, so that $\mathrm{d}s = \alpha\,\mathrm{d}u$, we have

$$\widehat{f_\lambda}(\xi) = \frac{1}{\pi^{\frac{d+1}{2}}\lambda^d} \int_0^\infty \left(\frac{s}{\alpha}\right)^{\frac{d-1}{2}} e^{-s} \frac{\mathrm{d}s}{\alpha} = \frac{1}{\pi^{\frac{d+1}{2}}\lambda^d\alpha^{\frac{d+1}{2}}} \int_0^\infty s^{\frac{d-1}{2}} e^{-s} \,\mathrm{d}s. \tag{21.128}$$

This is the integral of the Gamma function, specifically, $\Gamma\left(\frac{d+1}{2}\right)$. So we have

$$\widehat{f_\lambda}(\xi) = \frac{1}{\pi^{\frac{d+1}{2}}\lambda^d\alpha^{\frac{d+1}{2}}} \Gamma\left(\frac{d+1}{2}\right) = C_d \frac{1}{\lambda^d} \left(1 + \frac{|\xi|^2}{\lambda^2}\right)^{-\frac{d+1}{2}} = \frac{C_d\lambda}{\left(\lambda^2 + |\xi|^2\right)^{\frac{d+1}{2}}}. \tag{21.129}$$

■

Exercise 21.2. Let f be a function in $C_c^2\left(\mathbb{R}^d\right)$. Prove that

$$\sum_{i,j=1}^d \int_{\mathbb{R}^d} \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right|^2 \mathrm{d}x \le C \sum_{i=1}^d \int_{\mathbb{R}^d} \left| \frac{\partial^2}{\partial x_i \partial x_i} f(x) \right|^2 \mathrm{d}x. \tag{21.130}$$

What is the optimal (i.e. the smallest) constant C in this inequality?

Solution. We prove that $C = d$ is the optimal constant. Let $\widehat{f}$ be the Fourier transform.

$$\widehat{f}(\xi) = \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i x\cdot\xi}\, \mathrm{d}x. \tag{21.131}$$

Since $f \in C_c^2$, all second derivatives are in C_c, so they are in all L^p, and in particular, L^1. by Proposition 20.6(vi), we have

$$\widehat{\frac{\partial^2 f}{\partial x_i \partial x_j}}(\xi) = (2\pi i\xi_i)(2\pi i\xi_j)\,\widehat{f}(\xi) = -4\pi^2 \xi_i\xi_j \widehat{f}(\xi). \tag{21.132}$$

By the isometry property (Theorem 21.8), L^2-norm is preserved under Fourier transform. So we have

$$\int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|^2 \mathrm{d}x = \int_{\mathbb{R}^d} \left|\widehat{\frac{\partial^2 f(\xi)}{\partial x_i \partial x_j}}\right|^2 \mathrm{d}\xi = 16\pi^4 \int_{\mathbb{R}^d} \xi_i^2\xi_j^2 \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi. \tag{21.133}$$

Then the LHS of (21.130) reads

$$\mathrm{LHS} = \sum_{i,j=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|^2 \mathrm{d}x = 16\pi^4 \int_{\mathbb{R}^d} \left(\sum_{i,j=1}^d \xi_i^2\xi_j^2\right) \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi \tag{21.134}$$

$$= 16\pi^4 \int_{\mathbb{R}^d} \left(\sum_{i=1}^d \xi_i^2\right)^2 \left|\widehat{f}(\xi)\right| 2\mathrm{d}\xi. \tag{21.135}$$

By Cauchy–Schwarz inequality (the series version),

$$\left(\sum_{i=1}^d \xi_i^2\right)^2 \le \left(\sum_{i=1}^d (\xi_i^2)^2\right)\left(\sum_{i=1}^d 1^2\right) = d\sum_{i=1}^d \xi_i^4. \tag{21.136}$$

So we have

$$\mathrm{LHS} \le d16\pi^4 \int_{\mathbb{R}^d} \left(\sum_{i=1}^d \xi_i^4\right) \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi. \tag{21.137}$$

Now, in (21.133), if we take $i = j$, and sum over all i, then we would get

$$\sum_{i=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i^2}\right|^2 \mathrm{d}x = \sum_{i=1}^d \left[16\pi^4 \int_{\mathbb{R}^d} \xi_i^4 \left|\widehat{f}(\xi)\right| 2\mathrm{d}\xi\right] = 16\pi^4 \int_{\mathbb{R}^d} \left(\sum_{i=1}^d \xi_i^4\right) \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi. \tag{21.138}$$

So by combining (21.137) and (21.138), we have

$$\sum_{i,j=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|^2 \mathrm{d}x \le d\sum_{i=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i^2}\right|^2 \mathrm{d}x \tag{21.139}$$

To show that $C = d$ is the optimal constant, consider a smooth radial bump function $\phi \in C_c^\infty\left(\mathbb{R}^d\right)$, and take

$$f_\varepsilon(x) = \varepsilon^{\frac{d}{2}} \phi(\varepsilon x)\, e^{2\pi i x\cdot\xi_0}, \tag{21.140}$$

where $\xi_0 = (1, 1, \ldots, 1) \in \mathbb{R}^d$. Then $f_\varepsilon \in C_c^\infty \subseteq C_c^2$. Then the Fourier transform of f_ε is

$$\widehat{f_\varepsilon}(\xi) = \int_{\mathbb{R}^d} \varepsilon^{\frac{d}{2}} \phi(\varepsilon x)\, e^{2\pi i x\cdot\xi_0} e^{-2\pi i x\cdot\xi}\, \mathrm{d}x = \varepsilon^{\frac{d}{2}} \int_{\mathbb{R}^d} \phi(\varepsilon x)\, e^{-2\pi i x\cdot(\xi-\xi_0)}\, \mathrm{d}x. \tag{21.141}$$

Taking $y = \varepsilon x$, we have

$$\widehat{f_\varepsilon}(\xi) = \varepsilon^{\frac{d}{2}} \int_{\mathbb{R}^d} \phi(y)\, e^{-2\pi i y\cdot\frac{\xi-\xi_0}{\varepsilon}}\, \varepsilon^{-d}\, \mathrm{d}y = \varepsilon^{-\frac{d}{2}}\, \widehat{\phi}\left(\frac{\xi-\xi_0}{\varepsilon}\right). \tag{21.142}$$

Let

$$I = \sum_{i,j=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right|^2 \mathrm{d}x, \qquad J = \sum_{i=1}^d \int_{\mathbb{R}^d} \left|\frac{\partial^2 f(x)}{\partial x_i^2}\right|^2 \mathrm{d}x. \tag{21.143}$$

By (21.135) and (21.138), we have

$$\frac{I}{J} = \frac{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d \xi_i^2\right)^2 \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi}{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d \xi_i^4\right) \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi}. \tag{21.144}$$

For our chosen function f_ε, we do the change of variables $\xi = \xi_0 + \varepsilon\eta$. So we have

$$\left|\widehat{f_\varepsilon}(\xi)\right|^2 \mathrm{d}\xi = \varepsilon^{-d} \left|\widehat{\phi}(\eta)\right|^2\, \varepsilon^d\, \mathrm{d}\eta = \left|\widehat{\phi}(\eta)\right|^2\, \mathrm{d}\eta. \tag{21.145}$$

So we have

$$\frac{I}{J} = \frac{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d (1+\varepsilon\eta_i)^2\right)^2 \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta}{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d (1+\varepsilon\eta_i)^4\right) \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta}. \tag{21.146}$$

By Proposition 21.15, since $\phi \in C_c^\infty \subseteq \mathcal{S}(\mathbb{R}^d)$, we have $\widehat{\phi} \in \mathcal{S}(\mathbb{R}^d)$. By Lemma 21.14, since the integrands are just polynomials in ξ times a Schwartz function, they are in $\mathcal{S} \subseteq L^1$, as well.

In particular, the integrands are bouned by an integrable function, so we can apply Dominated Convergence Theorem to get

$$\lim_{\varepsilon\to 0} \frac{I}{J} = \frac{\int_{\mathbb{R}^d} \lim_{\varepsilon\to 0} \left(\sum_{i=1}^d (1+\varepsilon\eta_i)^2\right)^2 \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta}{\int_{\mathbb{R}^d} \lim_{\varepsilon\to 0} \left(\sum_{i=1}^d (1+\varepsilon\eta_i)^4\right) \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta} \tag{21.147}$$

$$= \frac{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d 1^2\right)^2 \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta}{\int_{\mathbb{R}^d} \left(\sum_{i=1}^d 1^4\right) \left|\widehat{\phi}(\eta)\right|^2 \mathrm{d}\eta} \tag{21.148}$$

$$= \frac{\left(\sum_{i=1}^d 1^2\right)^2}{\sum_{i=1}^d 1^4} = d. \tag{21.149}$$

Therefore, $\frac{I}{J}$ can get arbitrarily close to d, so d is the optimal constant. ■

Exercise 21.3. Let α, β be multi-indices and a, b real numbers. Suppose that for each $f \in$

$C_c^\infty\left(\mathbb{R}^d\right)$ one has

$$\left\| |\xi|^a \partial^\alpha \widehat{f} \right\|_{L^q\left(\mathbb{R}^d\right)} \leq C \left\| |x|^b \partial^\beta f \right\|_{L^p\left(\mathbb{R}^d\right)}. \tag{21.150}$$

What should be the relation between $a, b, \alpha, \beta, p, q, d$?

Solution. Let $f \in C_c^\infty\left(\mathbb{R}^d\right)$. If the inequality (21.150) holds for f, it must hold for the dialated function

$$f_\lambda(x) = f(\lambda x), \tag{21.151}$$

for any $\lambda > 0$. By chain rule,

$$\partial^\beta f_\lambda(x) = \lambda^{|\beta|} \partial^\beta f(\lambda x). \tag{21.152}$$

Therefore, the RHS of (21.150), computed for f_λ, is

$$\left\| |x|^b \partial^\beta f_\lambda \right\|_p^p = \int_{\mathbb{R}^d} \left[|x|^b \left| \partial^\beta f_\lambda(x) \right| \right]^p \mathrm{d}x = \int_{\mathbb{R}^d} |x|^{bp} \lambda^{|\beta| p} \left| \partial^\beta f(\lambda x) \right|^p \mathrm{d}x. \tag{21.153}$$

Doing the change of variables $y = \lambda x$, we have

$$\left\| |x|^b \partial^\beta f_\lambda \right\|_p^p = \lambda^{|\beta| p} \int_{\mathbb{R}^d} |y|^{bp} \lambda^{-bp} \left| \partial^\beta f(y) \right|^p \lambda^{-d} \mathrm{d}y = \lambda^{|\beta| p - bp - d} \left\| |y|^b \partial^\beta f \right\|_p^p. \tag{21.154}$$

So we have

$$\left\| |x|^b \partial^\beta f_\lambda \right\|_p = \lambda^{|\beta| - b - \frac{d}{p}} \left\| |y|^b \partial^\beta f \right\|_p. \tag{21.155}$$

Now let's compute the LHS of (21.150) for f_λ. We know that

$$\widehat{f_\lambda}(\xi) = \lambda^{-d} \widehat{f}\left(\frac{\xi}{\lambda}\right). \tag{21.156}$$

By chain rule,

$$\partial^\alpha \widehat{f_\lambda}(\xi) = \lambda^{-d-|\alpha|} \partial^\alpha \widehat{f}\left(\frac{\xi}{\lambda}\right). \tag{21.157}$$

Therefore,

$$\left\| |\xi|^a \partial^\alpha \widehat{f_\lambda} \right\|_q^q = \int_{\mathbb{R}^d} \left[|\xi|^a \left| \partial^\alpha \widehat{f_\lambda}(\xi) \right| \right]^q \mathrm{d}\xi = \int_{\mathbb{R}^d} |\xi|^{aq} \lambda^{-dq-|\alpha| q} \left| \partial^\alpha \widehat{f}\left(\frac{\xi}{\lambda}\right) \right|^q \mathrm{d}\xi. \tag{21.158}$$

We do the change of variables $\eta = \frac{\xi}{\lambda}$, we have

$$\left\| |\xi|^a \partial^\alpha \widehat{f_\lambda} \right\|_q^q = \lambda^{-dq-|\alpha| q} \int_{\mathbb{R}^d} |\eta|^{aq} \lambda^{aq} \left| \partial^\alpha \widehat{f}(\eta) \right|^q \lambda^d \mathrm{d}\xi = \lambda^{-dq-|\alpha| q+aq+d} \left\| |\eta|^a \partial^\alpha \widehat{f} \right\|_q^q. \tag{21.159}$$

So we have

$$\left\| |\xi|^a \partial^\alpha \widehat{f_\lambda} \right\|_q = \lambda^{-d-|\alpha|+a+\frac{d}{q}} \left\| |\eta|^a \partial^\alpha \widehat{f} \right\|_q. \tag{21.160}$$

Plugging things into (21.150), we get

$$\lambda^{-d-|\alpha|+a+\frac{d}{q}} \left\| |\eta|^a \partial^\alpha \widehat{f} \right\|_q \leq C \lambda^{|\beta|-b-\frac{d}{p}} \left\| |y|^b \partial^\beta f \right\|_p. \tag{21.161}$$

Since C is a constant independent of f, then we must have that the exponents of λ on both

sides should match (same reasoning as Remark 9.7). So we have

$$-d-|\alpha|+a+\frac{d}{q}=|\beta|-b-\frac{d}{p}. \tag{21.162}$$

Therefore, for an inequality of the form (21.150) to exist, the relation among $a, b, \alpha, \beta, p, q$ should be $-d-|\alpha|+a+\frac{d}{q}=|\beta|-b-\frac{d}{p}$. ■

Exercise 21.4. The purpose of this exercise is to show that $\mathcal{F}\left(L^1\left(\mathbb{R}^d\right)\right) \subsetneq C_0\left(\mathbb{R}^d\right)$.

(i) In dimension $d=1$, let $f \in L^1(\mathbb{R})$ and assume that $\widehat{f}$ is an odd function. Show that

$$\left|\int_1^a \frac{\widehat{f}(x)}{x} \mathrm{d}x\right| \leq C, \tag{21.163}$$

for any $a>1$, with C independent of a.

(ii) Find an odd function $g \in C_0(\mathbb{R})$ such that

$$\left|\int_1^a \frac{g(x)}{x} \mathrm{d}x\right| \tag{21.164}$$

is not bounded as $a \to \infty$.

Solution. (i) Since $\widehat{f}$ is an odd function, we can write it as

$$\widehat{f}(x)=\frac{\widehat{f}(x)-\widehat{f}(-x)}{2}=\int_{\mathbb{R}} f(t) \frac{e^{-2\pi i t x}-e^{2\pi i t x}}{2} \mathrm{d}t=-i \int_{\mathbb{R}} f(t) \sin(2\pi t x)\, \mathrm{d}t. \tag{21.165}$$

So we have

$$I:=\int_1^a \frac{\widehat{f}(x)}{x} \mathrm{d}x=-i \int_1^a\left[\int_{\mathbb{R}} f(t) \frac{\sin(2\pi t x)}{x} \mathrm{d}t\right] \mathrm{d}x. \tag{21.166}$$

We wish to apply Fubini's theorem to swap the integrals. The integrand is bounded (in absolute value) by $|f(t)|$, which is integrable on $(x,t) \in (1,a) \times \mathbb{R}$, and its integral is $(a-1)\|f\|_1$. Therefore, by Fubini's theorem,

$$I=-i \int_{\mathbb{R}} f(t)\left[\int_1^a \frac{\sin(2\pi t x)}{x} \mathrm{d}x\right] \mathrm{d}t. \tag{21.167}$$

In the inner integral, we perform the change of variables, $u=2\pi|t| x$. Then $\frac{\mathrm{d}x}{x}=\frac{\mathrm{d}u}{u}$, so the inner integral becomes

$$\int_1^a \frac{\sin(2\pi t x)}{x} \mathrm{d}x=\operatorname{sgn}(t) \int_{2\pi|t|}^{2\pi|t| a} \frac{\sin u}{u} \mathrm{d}u \tag{21.168}$$

$$=\operatorname{sgn}(t)\left[\operatorname{Si}(2\pi|t| a)-\operatorname{Si}(2\pi|t|)\right], \tag{21.169}$$

where $\operatorname{Si}(x)=\int_0^x \frac{\sin t}{t} \mathrm{d}t$ is the sine-integral function.

Claim 21.2: $\mathrm{Si} : \mathbb{R} \to \mathbb{R}$ is bounded.

Proof. Since $\lim_{|x|\to\infty} \mathrm{Si}(x) = \frac{\pi}{2}$, there exists some M such that

$$|\mathrm{Si}(x)| \leq \frac{\pi}{2} + 1 \text{ for } |x| > M. \tag{21.170}$$

Now, for $|x| \leq M$, we have

$$|\mathrm{Si}(x)| = |\mathrm{Si}(|x|)| = \left| \int_0^{|x|} \frac{\sin t}{t} \, \mathrm{d}t \right| \tag{21.171}$$

$$\leq \int_0^{|x|} \left| \frac{\sin t}{t} \right| \mathrm{d}t \leq \int_0^{|x|} 1 \, \mathrm{d}t = |x| \leq M. \tag{21.172}$$

Therefore, for all x, $|\mathrm{Si}(x)| \leq \max\left\{\frac{\pi}{2} + 1, M\right\}$. □

Therefore,

$$\left| \int_1^a \frac{\sin(2\pi t x)}{x} \, \mathrm{d}x \right| \leq 2M, \tag{21.173}$$

where $M = \sup_{x\in\mathbb{R}} |\mathrm{Si}(x)| < \infty$. So we have

$$|I| = \left| \int_{\mathbb{R}} f(t) \left[\int_1^a \frac{\sin(2\pi t x)}{x} \, \mathrm{d}x \right] \mathrm{d}t \right| \tag{21.174}$$

$$\leq \int_{\mathbb{R}} |f(t)| \left| \int_1^a \frac{\sin(2\pi t x)}{x} \, \mathrm{d}x \right| \mathrm{d}t \leq 2M \|f\|_1 . \tag{21.175}$$

(ii) We consider the function

$$g(x) = \begin{cases} \frac{x}{e} & \text{if } |x| < e, \\ \frac{\mathrm{sgn}(x)}{\log|x|} & \text{if } |x| \geq e. \end{cases} \tag{21.176}$$

This is a continuous, odd function, which decays to 0 when $|x| \to \infty$. Now,

$$\int_1^a \frac{g(x)}{x} \, \mathrm{d}x = \int_1^e \frac{1}{e} \, \mathrm{d}x + \int_e^a \frac{1}{x \log x} \, \mathrm{d}x \tag{21.177}$$

$$= \frac{e-1}{e} + [\log\log x]_e^a = \frac{e-1}{e} + \log\log a, \tag{21.178}$$

which is unbounded as $a \to \infty$.

■

Exercise 21.5. (i) Let $0 < \alpha < d$ and define $C_\alpha = \frac{\Gamma\left(\frac{\alpha}{2}\right)}{\pi^{\frac{\alpha}{2}}}$. Prove that the Fourier transform of $C_\alpha |x|^{-\alpha}$ is $C_{d-\alpha} |x|^{-d+\alpha}$ in the tempered distribution sense, i.e. for each φ in the Schwartz class $\mathcal{S}$ we have

$$\int_{\mathbb{R}^d} C_{d-\alpha} |x|^{-d+\alpha} \, \widehat{\varphi}(x) \, \mathrm{d}x = \int_{\mathbb{R}^d} C_\alpha |x|^{-\alpha} \, \varphi(x) \, \mathrm{d}x. \tag{21.179}$$

(ii) Compute explicitly $g(x) = |x|^{-\alpha} * |x|^{-\beta}$, where $0 < \alpha, \beta < d$ and $d < \alpha + \beta < 2d$.

Solution. (i) The integral formula for the Γ function is

$$\Gamma(\beta) = \int_0^\infty t^{\beta-1} e^{-t} \, \mathrm{d}t. \tag{21.180}$$

We do the change of variables $t = \pi\lambda |x|^2$ for $x \in \mathbb{R}^d - \{0\}$, so that $\mathrm{d}t = \pi |x|^2 \, \mathrm{d}\lambda$.

$$\Gamma(\beta) = \int_0^\infty \left(\pi\lambda |x|^2\right)^{\beta-1} e^{-\pi\lambda|x|^2} \pi |x|^2 \, \mathrm{d}\lambda = \pi^\beta |x|^{2\beta} \int_0^\infty \lambda^{\beta-1} e^{-\pi\lambda|x|^2} \, \mathrm{d}\lambda \tag{21.181}$$

Taking $\beta = \frac{\alpha}{2}$, we have

$$\frac{\Gamma\left(\frac{\alpha}{2}\right)}{\pi^{\frac{\alpha}{2}}} |x|^{-\alpha} = \int_0^\infty \lambda^{\frac{\alpha}{2}-1} e^{-\pi\lambda|x|^2} \, \mathrm{d}\lambda. \tag{21.182}$$

Then for any $\varphi \in \mathcal{S}\left(\mathbb{R}^d\right)$, we have

$$I := \int_{\mathbb{R}^d} C_\alpha |x|^{-\alpha} \varphi(x) \, \mathrm{d}x = \int_{\mathbb{R}^d} \left(\int_0^\infty \lambda^{\frac{\alpha}{2}-1} e^{-\pi\lambda|x|^2} \, \mathrm{d}\lambda \right) \varphi(x) \, \mathrm{d}x. \tag{21.183}$$

Claim 21.3: The integrand on the LHS of (21.183) is integrable.

Proof. For $|x| \le 1$, since $\varphi \in \mathcal{S}$, it is continuous, and hence, it's bounded by some constant M in this region. Switching to polar coordinates, we get

$$\int_{|x|\le 1} |x|^{-\alpha} |\varphi(x)| \, \mathrm{d}x \le M\omega_{d-1} \int_0^1 r^{-\alpha} r^{d-1} \, \mathrm{d}r = \left[\frac{r^{d-\alpha}}{d-\alpha} \right]_0^1 = \frac{1}{d-\alpha} < \infty. \tag{21.184}$$

On $|x| > 1$, $|x|^{-\alpha} < 1$, so we have

$$\int_{|x|>1} |x|^{-\alpha} |\varphi(x)| \, \mathrm{d}x \le \int_{|x|>1} |\varphi(x)| \, \mathrm{d}x = \|\varphi\|_1 < \infty. \tag{21.185}$$

So we have

$$\int_{\mathbb{R}^d} |x|^{-\alpha} |\varphi(x)| \, \mathrm{d}x \le \frac{1}{d-\alpha} + \|\varphi\|_1 < \infty. \tag{21.186}$$

□

Therefore, we can apply Fubini's theorem on (21.183).

$$I = \int_{\mathbb{R}^d} C_\alpha |x|^{-\alpha} \varphi(x) \, \mathrm{d}x = \int_0^\infty \lambda^{\frac{\alpha}{2}-1} \left(\int_{\mathbb{R}^d} e^{-\pi\lambda|x|^2} \varphi(x) \, \mathrm{d}x \right) \mathrm{d}\lambda. \tag{21.187}$$

Now, $e^{-\pi\lambda|x|^2}$ is the Fourier transform of $\lambda^{-\frac{d}{2}} e^{-\frac{\pi|x|^2}{\lambda}}$ by Proposition 21.2 (Fourier transform

of the Gaussian). Therefore, by Multiplication formula,

$$\int_{\mathbb{R}^d} e^{-\pi\lambda|x|^2}\,\varphi(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} \lambda^{-\frac{d}{2}} e^{-\frac{\pi|x|^2}{\lambda}}\,\widehat{\varphi}(x)\,\mathrm{d}x = \lambda^{-\frac{d}{2}} \int_{\mathbb{R}^d} e^{-\frac{\pi|x|^2}{\lambda}}\,\widehat{\varphi}(x)\,\mathrm{d}x. \tag{21.188}$$

So

$$I = \int_0^\infty \lambda^{\frac{\alpha}{2}-1}\lambda^{-\frac{d}{2}} \left(\int_{\mathbb{R}^d} e^{-\frac{\pi|x|^2}{\lambda}}\,\widehat{\varphi}(x)\,\mathrm{d}x\right)\mathrm{d}\lambda \tag{21.189}$$

$$= \int_0^\infty \lambda^{\frac{\alpha-d}{2}-1} \left(\int_{\mathbb{R}^d} e^{-\frac{\pi|x|^2}{\lambda}}\,\widehat{\varphi}(x)\,\mathrm{d}x\right)\mathrm{d}\lambda. \tag{21.190}$$

Claim 21.4: The integrand of (21.190) is integrable.

Proof. By Tonelli's theorem,

$$J := \int_0^\infty \int_{\mathbb{R}^d} \left|\lambda^{\frac{\alpha-d}{2}-1} e^{-\frac{\pi|x|^2}{\lambda}}\,\widehat{\varphi}(x)\right| \mathrm{d}x\,\mathrm{d}\lambda \tag{21.191}$$

$$= \int_{\mathbb{R}^d} |\widehat{\varphi}(x)| \int_0^\infty \left|\lambda^{\frac{\alpha-d}{2}-1} e^{-\frac{\pi|x|^2}{\lambda}}\right| \mathrm{d}\lambda\,\mathrm{d}x. \tag{21.192}$$

After the change of variables $\mu = \frac{1}{\lambda}$, we have $\mathrm{d}\mu = -\lambda^{-2}\,\mathrm{d}\lambda$, so we get

$$J = \int_{\mathbb{R}^d} |\widehat{\varphi}(x)| \int_0^\infty \mu^{-\frac{\alpha-d}{2}+1} e^{-\pi|x|^2\mu}\,\mu^{-2}\,\mathrm{d}\mu\,\mathrm{d}x \tag{21.193}$$

$$= \int_{\mathbb{R}^d} |\widehat{\varphi}(x)| \int_0^\infty \mu^{\frac{d-\alpha}{2}-1} e^{-\pi|x|^2\mu}\,\mathrm{d}\mu\,\mathrm{d}x \tag{21.194}$$

By (21.182), this is precisely $C_{d-\alpha}\,|x|^{-(d-\alpha)}$. So

$$J = \int_{\mathbb{R}^d} |\widehat{\varphi}(x)|\,C_{d-\alpha}\,|x|^{-(d-\alpha)}\,, \tag{21.195}$$

which is finite by the exact same computation as in Claim 21.3 since $\widehat{\varphi} \in \mathcal{S}$ (by Proposition 21.15). □

Therefore, we can apply Fubini's theorem on (21.190) to swap the order of integration. Then we have

$$I = \int_{\mathbb{R}^d} \widehat{\varphi}(x) \int_0^\infty \lambda^{\frac{\alpha-d}{2}-1} e^{-\frac{\pi|x|^2}{\lambda}}\,\mathrm{d}\lambda\,\mathrm{d}x. \tag{21.196}$$

The inner integral is exactly $C_{d-\alpha}\,|x|^{-(d-\alpha)}$, as shown in Claim 21.4. So we have

$$\int_{\mathbb{R}^d} C_\alpha\,|x|^{-\alpha}\,\varphi(x)\,\mathrm{d}x = I = \int_{\mathbb{R}^d} \widehat{\varphi}(x) C_{d-\alpha}\,|x|^{-(d-\alpha)}\,\mathrm{d}x, \tag{21.197}$$

as required.

(ii) We have

$$g(x) = \int_{\mathbb{R}^d} |x-y|^{-\alpha}\,|y|^{-\beta}\,\mathrm{d}y. \tag{21.198}$$

Using (21.182), one can write

$$|x-y|^{-\alpha} = \frac{1}{C_\alpha}\int_0^\infty s^{\frac{\alpha}{2}-1}e^{-\pi s|x-y|^2}\,\mathrm{d}s, \quad |y|^{-\beta} = \frac{1}{C_\beta}\int_0^\infty t^{\frac{\beta}{2}-1}e^{-\pi t|y|^2}\,\mathrm{d}t. \tag{21.199}$$

So $g(x)$ becomes a triple integral:

$$g(x) = \frac{1}{C_\alpha C_\beta}\int_{\mathbb{R}^d}\left(\int_0^\infty s^{\frac{\alpha}{2}-1}e^{-\pi s|x-y|^2}\,\mathrm{d}s\right)\left(\int_0^\infty t^{\frac{\beta}{2}-1}e^{-\pi t|y|^2}\,\mathrm{d}t.\right)\mathrm{d}y. \tag{21.200}$$

Everything is non-negative, so we can swap the order of integration as we please, by Tonelli's theorem.

$$g(x) = \frac{1}{C_\alpha C_\beta}\int_0^\infty\int_0^\infty s^{\frac{\alpha}{2}-1}t^{\frac{\beta}{2}-1}\left(\int_{\mathbb{R}^d} e^{-\pi s|x-y|^2-\pi t|y|^2}\,\mathrm{d}y\right)\mathrm{d}s\,\mathrm{d}t. \tag{21.201}$$

We complete the squares in the inner integral:

$$\begin{aligned} s\,|x-y|^2 + t\,|y|^2 &= s\,|x|^2 + s\,|y|^2 - 2s\,(x\cdot y) + t\,|y|^2 \\ &= (s+t)\,|y|^2 - 2s\,(x\cdot y) + s\,|x|^2 \\ &= (s+t)\left[|y|^2 - \frac{2s}{s+t}\,(x\cdot y) + \left(\frac{s}{s+t}\right)|x|^2\right] - \frac{s^2}{s+t}\,|x|^2 + s\,|x|^2 \\ &= (s+t)\left|y - \frac{s}{s+t}x\right|^2 + \frac{st}{s+t}\,|x|^2\,. \end{aligned} \tag{21.202}$$

So the inner integral becomes

$$\int_{\mathbb{R}^d}\exp\left(-\pi(s+t)\left|y-\frac{s}{s+t}x\right|^2\right)\exp\left(-\frac{\pi st}{s+t}\,|x|^2\right)\mathrm{d}y \tag{21.203}$$

Since $\int_{\mathbb{R}} e^{-\pi a x^2} = \frac{1}{\sqrt{a}}$, the integral in (21.203) is actually

$$\int_{\mathbb{R}^d} e^{-\pi s|x-y|^2-\pi t|y|^2}\,\mathrm{d}y = e^{-\frac{\pi st}{s+t}|x|^2}\,(s+t)^{-\frac{d}{2}}\,. \tag{21.204}$$

Hence,

$$g(x) = \frac{1}{C_\alpha C_\beta}\int_0^\infty\int_0^\infty s^{\frac{\alpha}{2}-1}t^{\frac{\beta}{2}-1}e^{-\frac{\pi st}{s+t}|x|^2}\,(s+t)^{-\frac{d}{2}}\,\mathrm{d}s\,\mathrm{d}t. \tag{21.205}$$

We do the change of variables $s = ur$ and $t = u(1-r)$, with $u \in (0,\infty)$ and $r \in (0,1)$. Then the Jacobian determinant is

$$\left|\frac{\partial\,(s,t)}{\partial\,(u,r)}\right| = \begin{vmatrix} r & u \\ 1-r & -u \end{vmatrix} = u. \tag{21.206}$$

Note also that $s+t=u$ and $\frac{st}{s+t}=ur(1-r)$. Then

$$g(x)=\frac{1}{C_\alpha C_\beta}\int_0^1\int_0^\infty (ur)^{\frac{\alpha}{2}-1}(u(1-r))^{\frac{\beta}{2}-1}e^{-\pi ur(1-r)|x|^2}u^{-\frac{d}{2}}\,u\,\mathrm{d}u\,\mathrm{d}r \tag{21.207}$$

$$=\frac{1}{C_\alpha C_\beta}\int_0^1 r^{\frac{\alpha}{2}-1}(1-r)^{\frac{\beta}{2}-1}\int_0^\infty u^{\frac{\alpha+\beta-d}{2}-1}e^{-\pi ur(1-r)|x|^2}\,\mathrm{d}u\,\mathrm{d}r \tag{21.208}$$

By (21.182), the inner integral is

$$C_{\alpha+\beta-d}\left(r(1-r)\,|x|^2\right)^{-\frac{\alpha+\beta-d}{2}}. \tag{21.209}$$

Plugging it back to the expression of $g(x)$, we have

$$g(x)=\frac{C_{\alpha+\beta-d}}{C_\alpha C_\beta}\,|x|^{-(\alpha+\beta-d)}\int_0^1 r^{\frac{d-\beta}{2}-1}(1-r)^{\frac{d-\alpha}{2}-1}\,\mathrm{d}r. \tag{21.210}$$

This is the integral of Beta function. The integral is

$$B\left(\frac{d-\beta}{2},\frac{d-\alpha}{2}\right)=\frac{\Gamma\left(\frac{d-\alpha}{2}\right)\Gamma\left(\frac{d-\beta}{2}\right)}{\Gamma\left(\frac{2d-\alpha-\beta}{2}\right)}. \tag{21.211}$$

Combining everything,

$$g\left(x\right)=\frac{C_{\alpha+\beta-d}}{C_\alpha C_\beta}\,|x|^{-(\alpha+\beta-d)}\,\frac{\Gamma\left(\frac{d-\alpha}{2}\right)\Gamma\left(\frac{d-\beta}{2}\right)}{\Gamma\left(\frac{2d-\alpha-\beta}{2}\right)}. \tag{21.212}$$

Note that

$$\frac{C_{\alpha+\beta-d}}{C_\alpha C_\beta}=\frac{\frac{\Gamma\left(\frac{\alpha+\beta-d}{2}\right)}{\pi^{\frac{\alpha+\beta-d}{2}}}}{\frac{\Gamma\left(\frac{\alpha}{2}\right)}{\pi^{\frac{\alpha}{2}}}\frac{\Gamma\left(\frac{\beta}{2}\right)}{\pi^{\frac{\beta}{2}}}}=\pi^{\frac{d}{2}}\frac{\Gamma\left(\frac{\alpha+\beta-d}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{\beta}{2}\right)}. \tag{21.213}$$

So

$$g\left(x\right)=\pi^{\frac{d}{2}}\,\frac{\Gamma\left(\frac{\alpha+\beta-d}{2}\right)\Gamma\left(\frac{d-\alpha}{2}\right)\Gamma\left(\frac{d-\beta}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{\beta}{2}\right)\Gamma\left(\frac{2d-\alpha-\beta}{2}\right)}\,|x|^{-(\alpha+\beta-d)}. \tag{21.214}$$

■

Exercise 21.6. Let $\operatorname{sinc}x=\frac{\sin(\pi x)}{\pi x}$, with $\operatorname{sinc}0=1$.

(i) If $a>0$, verify that

$$\left(\chi_{[-a,a]}\right)^\wedge(\xi)=\left(\chi_{[-a,a]}\right)^\vee(\xi)=2a\operatorname{sinc}\left(2a\xi\right). \tag{21.215}$$

(ii) Let $\mathcal{H}_a=\left\{f\in L^2\left(\mathbb{R}\right):\widehat{f}\left(\xi\right)=0\text{ (a.e.) for }|\xi|>a\right\}$. Show that $\mathcal{H}_a$ is a Hilbert space and that $\left\{\sqrt{2a}\operatorname{sinc}\left(2ax-k\right):k\in\mathbb{Z}\right\}$ is an orthonormal basis of $\mathcal{H}_a$.

(iii) Show that if $f \in \mathcal{H}_a$, then $f \in C_0(\mathbb{R})$ (possibly after a modification on a set of measure zero) and

$$f(x) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{2a}\right) \operatorname{sinc}(2ax - k), \tag{21.216}$$

where this series converges uniformly (and also in L^2).

Solution. (i) For $\xi = 0$,

$$\left(\chi_{[-a,a]}\right)^{\wedge}(0) = \int_{\mathbb{R}} e^{-2\pi i x \cdot 0} \chi_{[-a,a]}(x)\, \mathrm{d}x = \int_{-a}^{a} 1 \mathrm{d}x = 2a = 2a \operatorname{sinc}(0). \tag{21.217}$$

For $\xi \neq 0$, For $\xi = 0$,

$$\left(\chi_{[-a,a]}\right)^{\wedge}(\xi) = \int_{\mathbb{R}} e^{-2\pi i x \xi} \chi_{[-a,a]}(x)\, \mathrm{d}x = \int_{-a}^{a} e^{-2\pi i x \xi}\, \mathrm{d}x \tag{21.218}$$

$$= \left[\frac{e^{-2\pi i x \xi}}{-2\pi i \xi}\right]_{-a}^{a} = \frac{e^{-2\pi i a \xi} - e^{2\pi i a \xi}}{-2\pi i \xi} \tag{21.219}$$

$$= 2a \frac{\sin(2\pi a \xi)}{2\pi a \xi} = 2a \operatorname{sinc}(2a\xi). \tag{21.220}$$

By Plancherel's theorem,

$$\left(\chi_{[-a,a]}\right)^{\vee}(\xi) = \left(\chi_{[-a,a]}\right)^{\wedge}(-\xi) = 2a \operatorname{sinc}(2a\xi). \tag{21.221}$$

(ii) The Fourier transform $\mathcal{F} : L^2 \to L^2$ is an unitary isometry. Then $\mathcal{H}_a$ is the preimage of

$$V_a = \left\{ f \in L^2 \ : \ f(x) = 0 \text{ a.e. for } |x| > a \right\} \tag{21.222}$$

under $\mathcal{F}$.

Claim 21.5: $V_a \subseteq L^2$ is a closed subspace.

Proof. First V_a is a subspace, since if g_1, g_2 vanish a.e. outside $[-a, a]$, then $ag_1 + bg_2$ vanishes a.e. outside $[-a, a]$ (since union of nullsets is also a nullset). Now let $g_n \in V_a$ be a sequence that converges to $g \in L^2$. Now,

$$\|g_n - g\|_2^2 = \int_{\mathbb{R}} |g_n - g|^2 \geq \int_{|x|>a} |0 - g(x)|^2\, \mathrm{d}x = \int_{|x|>a} |g(x)|^2\, \mathrm{d}x. \tag{21.223}$$

If $n \to \infty$, the LHS goes to 0. So we have

$$\int_{|x|>a} |g(x)|^2\, \mathrm{d}x = 0. \tag{21.224}$$

So $g = 0$ a.e. on $|x| > a$. In other words, $g \in V_a$, proving V_a closed. □

Therefore, $\mathcal{H}_a = \mathcal{F}^{-1}(V_a)$ is a Hilbert space (since it's the preimage of a closed set under a continuous map, it's closed; and closed subspace of a complete space is complete). Now, V_a

can be (canonically) identified to $L^2([-a,a])$.

Claim 21.6: $\mathcal{B} = \left\{e^{2\pi inx} : n \in \mathbb{Z}\right\}$ is an orthonormal basis of $L^2([0,1])$.

Proof. Let $f_n(x) = e^{2\pi inx}$. Then

$$\langle f_n, f_m\rangle = \int_0^1 e^{2\pi inx} e^{-2\pi imx}\,\mathrm{d}x = \int_0^1 e^{2\pi i(n-m)x}\,\mathrm{d}x, \tag{21.225}$$

which is 1 for $n = m$. For $n \neq m$, we have

$$\langle f_n, f_m\rangle = \left[\frac{e^{2\pi i(n-m)x}}{2\pi i(n-m)}\right]_0^1 = 0. \tag{21.226}$$

So it's an orthonormal collection.

Let $\mathcal{A} = \operatorname{span}\mathcal{B}$, which is a subalgebra of $C([0,1])$ (which is dense in $L^2[0,1]$), which contains the constant function 1 (because $1 = f_0$), and for $x \neq y$, there exists some $f \in \mathcal{A}$, namely f_1, such that $f_1(x) \neq f_1(y)$. Therefore, by the Generaized Stone–Weierstrass theorem, $\mathcal{A}$ is dense in $C([0,1])$, so $\mathcal{A}$ is dense in $L^2[0,1]$. □

Therefore, on $L^2[-a,a]$,

$$e_k(x) := e^{-2\pi i\frac{k}{2a}x}\chi_{[-a,a]}(x), \quad k \in \mathbb{Z}, \tag{21.227}$$

is an orthogonal basis. To make it orthonormal, we divide by its norm $\sqrt{2a}$.

$$e_k(x) := \frac{1}{\sqrt{2a}} e^{-2\pi i\frac{k}{2a}x}\chi_{[-a,a]}(x), \quad k \in \mathbb{Z}. \tag{21.228}$$

Since $\mathcal{F}$ is a unitary isometry, preimages of an orthonormal basis of V_a is an orthonormal basis for $\mathcal{H}_a$. Now,

$$\mathcal{F}^{-1}e_k(x) = \frac{1}{\sqrt{2a}}\int_{-a}^{a} e^{-2\pi i\frac{k}{2a}\xi} e^{2\pi i\xi x}\,\mathrm{d}\xi = \frac{1}{\sqrt{2a}}\int_{-a}^{a} e^{-2\pi i\frac{k}{2a}\xi} e^{2\pi i\xi x}\,\mathrm{d}\xi \tag{21.229}$$

$$= \frac{1}{\sqrt{2a}}\int_{-a}^{a} e^{2\pi i\xi\left(x-\frac{k}{2a}\right)}\,\mathrm{d}\xi = \frac{1}{\sqrt{2a}}\left(\chi_{[-a,a]}\right)^{\vee}\left(x-\frac{k}{2a}\right) \tag{21.230}$$

$$= \frac{1}{\sqrt{2a}}\,2a\operatorname{sinc}\left(2a\left(x-\frac{k}{2a}\right)\right) = \sqrt{2a}\operatorname{sinc}(2ax-k). \tag{21.231}$$

Therefore, $\left\{\sqrt{2a}\operatorname{sinc}(2ax-k) : k \in \mathbb{Z}\right\}$ is an orthonormal basis for $\mathcal{H}_a$.

(iii) Since $f \in L^2$, we have $\widehat{f} \in L^2$. Now,

$$\left\|\widehat{f}\right\|_1 = \int_{-a}^{a}\left|\widehat{f}(\xi)\right|\,\mathrm{d}\xi \leq \|1\|_2\left\|\widehat{f}\right\|_2 = \sqrt{2a}\left\|\widehat{f}\right\|_2 < \infty, \tag{21.232}$$

so $\widehat{f} \in L^1 \cap L^2$. Now,

$$f(x) = \mathcal{F}^{-1}\widehat{f}(x) = \mathcal{F}\widehat{f}(-x) = \int_{-a}^{a} \widehat{f}(\xi)\, e^{2\pi i x\xi}\, \mathrm{d}\xi. \tag{21.233}$$

By Riemann–Lebesgue lemma, $\left(\widehat{f}\right)^{\vee}$ is in C_0. But $\left(\widehat{f}\right)^{\vee} = f$ a.e. So we have $f \in C_0$.

Now, given $f \in \mathcal{H}_a$, since

$$\psi_k(x) = \sqrt{2a}\,\mathrm{sinc}\,(2ax - k), \quad k \in \mathbb{Z} \tag{21.234}$$

is an orthonormal basis, we have that

$$f = \sum_{k\in\mathbb{Z}} \langle f, \psi_k\rangle\, \psi_k, \tag{21.235}$$

where this convergence is in L^2. Now,

$$\langle f, \psi_k\rangle = \left\langle \widehat{f}, e_k \right\rangle = \int_{-a}^{a} \widehat{f}(\xi)\, \overline{\frac{1}{\sqrt{2a}} e^{-2\pi i \frac{k}{2a}\xi}}\, \mathrm{d}\xi \tag{21.236}$$

$$= \frac{1}{\sqrt{2a}} \int_{-a}^{a} \widehat{f}(\xi)\, e^{2\pi i \frac{k}{2a}\xi}\, \mathrm{d}\xi = \frac{1}{\sqrt{2a}} f\left(\frac{k}{2a}\right). \tag{21.237}$$

Therefore,

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \psi_k\rangle\, \psi_k(x) = \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{2a}} f\left(\frac{k}{2a}\right) \sqrt{2a}\,\mathrm{sinc}\,(2ax - k) \tag{21.238}$$

$$= \sum_{k\in\mathbb{Z}} f\left(\frac{k}{2a}\right) \mathrm{sinc}\,(2ax - k), \tag{21.239}$$

where this convergence is in L^2. To show that it's an uniform convergence, let

$$S_N(x) = \sum_{k=-N}^{N} f\left(\frac{k}{2a}\right) \mathrm{sinc}\,(2ax - k). \tag{21.240}$$

Now, from (21.230) and (21.231),

$$\mathrm{sinc}\,(2ax - k) = \frac{1}{2a} \int_{-a}^{a} e^{2\pi i \xi\left(x - \frac{k}{2a}\right)}\, \mathrm{d}\xi, \tag{21.241}$$

and using the Fourier inversion formula $f(x) = \int_{-a}^{a} \widehat{f}(\xi)\, e^{2\pi i x\xi}\, \mathrm{d}\xi$, we have

$$f(x) - S_N(x) = \int_{-a}^{a} \widehat{f}(\xi)\, e^{2\pi i x\xi}\, \mathrm{d}\xi - \sum_{k=-N}^{N} f\left(\frac{k}{2a}\right) \frac{1}{2a} \int_{-a}^{a} e^{2\pi i \xi\left(x - \frac{k}{2a}\right)}\, \mathrm{d}\xi \tag{21.242}$$

$$= \int_{-a}^{a} \left[\widehat{f}(\xi) - \sum_{k=-N}^{N} f\left(\frac{k}{2a}\right) \frac{1}{2a} e^{-2\pi i \xi \frac{k}{2a}}\right] e^{2\pi i x\xi}\, \mathrm{d}\xi. \tag{21.243}$$

Now, from (21.237),

$$f\left(\frac{k}{2a}\right)\frac{1}{2a}=\frac{1}{\sqrt{2a}}\left\langle \widehat{f},e_k\right\rangle, \tag{21.244}$$

so we have

$$f\left(\frac{k}{2a}\right)\frac{1}{2a}e^{-2\pi i\xi\frac{k}{2a}}=\frac{1}{\sqrt{2a}}\left\langle \widehat{f},e_k\right\rangle e^{-2\pi i\xi\frac{k}{2a}}=\left\langle \widehat{f},e_k\right\rangle e_k\left(\xi\right). \tag{21.245}$$

Hence, the term inside the bracket of (21.243) is

$$\widehat{f}\left(\xi\right)-\sum_{k=-N}^{N}\left\langle \widehat{f},e_k\right\rangle e_k\left(\xi\right), \tag{21.246}$$

which converges to 0 in L^2 as $N\to\infty$ (since e_k is an orthonormal basis on $L^2\left[-a,a\right]$). Therefore, from (21.243),

$$\begin{aligned}\left|f\left(x\right)-S_N(x)\right| &\leq \int_{-a}^{a}\left|\widehat{f}-\sum_{k=-N}^{N}\left\langle \widehat{f},e_k\right\rangle e_k\right|\mathrm{d}\xi && (21.247)\\ &\leq \left\|\widehat{f}-\sum_{k=-N}^{N}\left\langle \widehat{f},e_k\right\rangle e_k\right\|_2\left\|1\right\|_2 && (21.248)\\ &=\sqrt{2a}\left\|\widehat{f}-\sum_{k=-N}^{N}\left\langle \widehat{f},e_k\right\rangle e_k\right\|_2. && (21.249)\end{aligned}$$

The RHS is independent of x, so

$$\sup_{x\in\mathbb{R}}\left|f\left(x\right)-S_N(x)\right|\leq\sqrt{2a}\left\|\widehat{f}-\sum_{k=-N}^{N}\left\langle \widehat{f},e_k\right\rangle e_k\right\|_2, \tag{21.250}$$

and the RHS approaches 0 as $N\to\infty$. So $S_N\to f$ uniformly.

■

Exercise 21.7 (Heisenberg's uncertainty principle)**.** Suppose that $f\in L^2\left(\mathbb{R}\right)$.

(i) Show that the L^2-derivative of f exists if and only if $\xi\widehat{f}\in L^2$, and in this case $\widehat{f'}\left(\xi\right)=2\pi i\xi\widehat{f}\left(\xi\right)$.

(ii) Show that if the L^2-derivative f' exists, then

$$\left(\int_{\mathbb{R}}\left|f\left(x\right)\right|^2\mathrm{d}x\right)^2\leq 4\int_{\mathbb{R}}\left|xf\left(x\right)\right|^2\mathrm{d}x\int_{\mathbb{R}}\left|f'\left(x\right)\right|^2\mathrm{d}x. \tag{21.251}$$

(iii) Conclude that for any $x_0,\xi_0\in\mathbb{R}$ we have

$$\int_{\mathbb{R}}\left(x-x_0\right)^2\left|f\left(x\right)\right|^2\mathrm{d}x\int_{\mathbb{R}}\left(\xi-\xi_0\right)^2\left|\widehat{f}\left(\xi\right)\right|^2\mathrm{d}\xi\geq\frac{\left\|f\right\|_{L^2}^4}{16\pi^2}. \tag{21.252}$$

That means that f and $\widehat{f}$ cannot be simultaneously localized around x_0 and ξ_0, respectively.

Solution. (i) Since Fourier transform preserves the L^2-norm, $D_h f(x) = \frac{f(x+h)-f(x)}{h}$ converges as $h \to 0$ if and only if $\widehat{D_h f}$ converges as $h \to 0$.

$$\widehat{D_h f} = \frac{\widehat{\tau_{-h} f - f}}{h} = \frac{e^{2\pi i h\xi} - 1}{h} \widehat{f}. \tag{21.253}$$

Suppose $\xi \widehat{f} \in L^2$. Then we claim that $\widehat{D_h f}$ converges to $2\pi i \xi \widehat{f}$ as $h \to 0$.

$$I(h) := \left| \widehat{D_h f} - 2\pi i \xi \widehat{f} \right|^2 = \left| \frac{e^{2\pi i h\xi} - 1}{h} - 2\pi i \xi \right|^2 \left| \widehat{f} \right|^2. \tag{21.254}$$

Claim 21.7: $\left| e^{ix} - 1 \right| \le |x|$ for all x.

Proof.

$$\left| e^{ix} - 1 \right|^2 = |\cos x + i \sin x - 1|^2 = (\cos x - 1)^2 + \sin^2 x \tag{21.255}$$

$$= 2 - 2\cos x = 4 \sin^2 \left(\frac{x}{2} \right). \tag{21.256}$$

Therefore,

$$\left| e^{ix} - 1 \right| \le |x| \iff 2 \left| \sin \left(\frac{x}{2} \right) \right| \le |x| \iff \left| \sin \left(\frac{x}{2} \right) \right| \le \left| \frac{x}{2} \right|, \tag{21.257}$$

which holds for all $x \in \mathbb{R}$. □

Using the fact that $\left| \frac{e^{ix}-1}{x} \right| \le 1$, we have

$$I(h) \le \left(\left| \frac{e^{2\pi i h\xi} - 1}{h} \right| + |2\pi i \xi| \right)^2 \left| \widehat{f} \right|^2 \le 16\pi^2 \left| \xi \widehat{f} \right|^2. \tag{21.258}$$

So we have

$$\lim_{h \to 0} \left\| \widehat{D_h f} - 2\pi i \xi \widehat{f} \right\|_2^2 = \lim_{h \to 0} \int_{\mathbb{R}} I(h) \, \mathrm{d}\xi, \tag{21.259}$$

and the integrand is bounded by the L^1 function $16\pi^2 \left| \xi \widehat{f} \right|^2$. So we can take the limit inside by Dominated Convergence Theorem to get

$$\lim_{h \to 0} \left\| \widehat{D_h f} - 2\pi i \xi \widehat{f} \right\|_2^2 = \int_{\mathbb{R}} \lim_{h \to 0} I(h) \, \mathrm{d}\xi = \int_{\mathbb{R}} \lim_{h \to 0} \left| \frac{e^{2\pi i h\xi} - 1}{h} - 2\pi i \xi \right|^2 \left| \widehat{f} \right|^2 \mathrm{d}\xi \tag{21.260}$$

$\frac{e^{2\pi i h\xi}-1}{h} \to 2\pi i \xi$ as $h \to 0$, we we have

$$\lim_{h \to 0} \left\| \widehat{D_h f} - 2\pi i \xi \widehat{f} \right\|_2^2 = \int_{\mathbb{R}} \lim_{h \to 0} \left| \frac{e^{2\pi i h\xi} - 1}{h} - 2\pi i \xi \right|^2 \left| \widehat{f} \right|^2 \mathrm{d}\xi = 0. \tag{21.261}$$

So $\widehat{D_h f}$ converges to some $g = 2\pi i \xi \widehat{f} \in L^2$. Since $\mathcal{F}$ is a surjective isometry on L^2, $g = \widehat{f'}$ for some $f' \in L^2$. By isometry, $D_h f$ converges in L^2 to f'. Then $\widehat{D_h f}$ converges to $\widehat{f'}$, which is $2\pi i \xi \widehat{f}$.

Now, conversely, suppose L^2-derivative of f exists. So $\widehat{D_h f}$ converge to some $\widehat{f'}$, and this convergence is in L^2. Therefore, there exists some subsequence $h_n \to 0$ such that for a.e. ξ,

$$\widehat{D_{h_n} f}(\xi) = \frac{e^{2\pi i h_n \xi} - 1}{h_n} \widehat{f} \to \widehat{f'}(\xi), \text{ as } n \to \infty. \tag{21.262}$$

As $h_n \to 0$, we have

$$\lim_{n \to \infty} \frac{e^{2\pi i h_n \xi} - 1}{h_n} = 2\pi i \xi. \tag{21.263}$$

Therefore, for a.e. ξ,

$$\widehat{f'}(\xi) = 2\pi i \xi \widehat{f}(\xi). \tag{21.264}$$

Now, $f' \in L^2$, so $\widehat{f'} \in L^2$. Therefore, $\xi \widehat{f}(\xi) \in L^2$.

(ii) We can assume $xf(x) \in L^2$, because otherwise the RHS is ∞ and we don't have anything to prove. Since $f' \in L^2$, by Cauchy–Schwarz inequality (or Hölder's inequality), $xf\overline{f'} \in L^1$.

Since $xf \in L^2$ and $D_h f \to f'$ in L^2 as $h \to 0$, we have

$$\lim_{h \to 0} \langle xf, D_h f \rangle = \langle xf, f' \rangle. \tag{21.265}$$

Now, by $|a-b|^2 = |a|^2 + |b|^2 - 2\operatorname{Re}(a\bar{b})$, we have

$$2\operatorname{Re}\left(f(x)\overline{f(x+h)}\right) = |f(x)|^2 + |f(x+h)|^2 - |f(x+h) - f(x)|^2. \tag{21.266}$$

Then we multiply by x and integrate. $x|f|^2 \in L^1$ by Hölder's inequality on $xf, \overline{f} \in L^2$. For $x|f(x+h)|^2$, we have

$$\int_{\mathbb{R}} x|f(x+h)|^2 \, \mathrm{d}x = \int_{\mathbb{R}} (x-h)|f(x)|^2 \, \mathrm{d}x = \int_{\mathbb{R}} x|f|^2 - h\|f\|_2^2. \tag{21.267}$$

Therefore,

$$2\operatorname{Re}\int_{\mathbb{R}} x\left(f(x)\overline{f(x+h)}\right) \mathrm{d}x = 2\int_{\mathbb{R}} x|f|^2 - h\|f\|_2^2 - \int_{\mathbb{R}} x|f(x+h) - f(x)|^2 \, \mathrm{d}x. \tag{21.268}$$

Subtracting $2\int_{\mathbb{R}} x|f|^2$ and dividing by h, we have

$$2\operatorname{Re}\langle xf, D_h f \rangle = -\|f\|_2^2 - \frac{1}{h}\int_{\mathbb{R}} x|f(x+h) - f(x)|^2 \, \mathrm{d}x. \tag{21.269}$$

We claim that as $h \to 0$, the last integral vanishes.

$$\begin{aligned}
&\left|\frac{1}{h}\int_{\mathbb{R}} x\left|f\left(x+h\right)-f\left(x\right)\right|^2 \mathrm{d}x\right| \\
=& \int_{\mathbb{R}} x\left(f\left(x+h\right)-f\left(x\right)\right)\overline{D_h f(x)}\,\mathrm{d}x \\
\leq& \left\|x\left(f\left(x+h\right)-f\left(x\right)\right)\right\|_2 \left\|D_h f(x)\right\|_2 \\
\leq& \left[\left\|(x+h)f\left(x+h\right)-xf\left(x\right)\right\|_2+\left\|hf(x+h)\right\|_2\right]\left\|D_h f(x)\right\|_2 . \qquad (21.270)
\end{aligned}$$

By Continuity of translations,

$$\lim_{h\to 0}\left\|(x+h)f\left(x+h\right)-xf\left(x\right)\right\|_2 = 0. \qquad (21.271)$$

Furthermore, $\|hf(x+h)\| = |h|\left\|f\right\|_2 \to 0$. And $\|D_h f(x)\|_2 \to \|f'\|_2 < \infty$. Therefore, the RHS of (21.270) vanishes as $h \to 0$. So we have

$$2\,\mathrm{Re}\left\langle xf, f'\right\rangle = 2\,\mathrm{Re}\lim_{h\to 0}\left\langle xf, D_h f\right\rangle = -\left\|f\right\|_2^2 . \qquad (21.272)$$

Now, by Cauchy–Schwarz inequality,

$$\left\|f\right\|_2^2 \leq 2\left|\int_{\mathbb{R}} x\,f\overline{f'}\right| \leq 2\left\|xf\right\|_2\left\|f'\right\|_2 . \qquad (21.273)$$

Squaring, we have

$$\left\|f\right\|_2^4 \leq 4\left\|xf\right\|_2^2\left\|f'\right\|_2^2 , \qquad (21.274)$$

as desired.

(iii) Let's define

$$h\left(x\right) = e^{-2\pi i\xi_0 x} f\left(x+x_0\right). \qquad (21.275)$$

Then $h \in L^2$, with $\left\|h\right\|_2 = \left\|f\right\|_2$.

Claim 21.8: $\widehat{h}\left(\xi\right) = e^{2\pi i x_0(\xi+\xi_0)}\widehat{f}\left(\xi+\xi_0\right)$.

Proof. Let f_n be a sequence in $L^1 \cap L^2$ converging to f in L^2. We define

$$h_n\left(x\right) = e^{-2\pi i\xi_0 x} f_n\left(x+x_0\right) \in L^1 \cap L^2. \qquad (21.276)$$

Then its Fourier transform is

$$\widehat{h}_n(\xi) = \int_{\mathbb{R}} e^{-2\pi i\xi_0 x} f_n\left(x+x_0\right)\ e^{-2\pi i x\xi}\,\mathrm{d}x = \int_{\mathbb{R}} e^{-2\pi i(\xi_0+\xi)x} f_n\left(x+x_0\right)\,\mathrm{d}x. \qquad (21.277)$$

By a change of variables $x + x_0 \to x$,

$$\widehat{h}_n(\xi) = \int_{\mathbb{R}} e^{-2\pi i(\xi_0+\xi)(x-x_0)} f_n\left(x\right)\,\mathrm{d}x. = e^{2\pi i x_0(\xi+\xi_0)}\widehat{f_n}\left(\xi+\xi_0\right). \qquad (21.278)$$

Since $\widehat{f_n}$ converges to $\widehat{f}$ in L^2, $\widehat{h_n}$ converges to $e^{2\pi i x_0(\xi+\xi_0)}\widehat{f}(\xi+\xi_0)$ in L^2. □

So we have

$$\widehat{h}(\xi) = e^{2\pi i x_0(\xi+\xi_0)}\widehat{f}(\xi+\xi_0). \tag{21.279}$$

Then $\left|\widehat{h}(\xi)\right| = \left|\widehat{f}(\xi+\xi_0)\right|$. Now we define

$$A := \int_{\mathbb{R}} (x-x_0)^2 |f(x)|^2 \, \mathrm{d}x = \int_{\mathbb{R}} x^2 |f(x+x_0)|^2 \, \mathrm{d}x = \int_{\mathbb{R}} x^2 |h(x)|^2 \, \mathrm{d}x \tag{21.280}$$

$$B := \int_{\mathbb{R}} (\xi-\xi_0)^2 \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi = \int_{\mathbb{R}} \xi^2 \left|\widehat{f}(\xi+\xi_0)\right|^2 \mathrm{d}\xi = \int_{\mathbb{R}} \xi^2 \left|\widehat{h}(\xi)\right|^2 \mathrm{d}\xi. \tag{21.281}$$

If $A=\infty$ or $B=\infty$, the inequality is trivial. So we assume both are finite. So $xh \in L^2$, and $\xi\widehat{h} \in L^2$. So by (i), $h' \in L^2$ exists with

$$\widehat{h'} = 2\pi i \xi \widehat{h}. \tag{21.282}$$

Then by isometry,

$$\|h'\|_2 = \left\|\widehat{h'}\right\|_2 = 4\pi^2 \int_{\mathbb{R}} \xi^2 \left|\widehat{h}\right|^2 \mathrm{d}\xi = 4\pi^2 B. \tag{21.283}$$

Now, applying (ii) to h, we have

$$\|f\|_2^4 = \|h\|_2^4 \le 4\|xh\|_2^2 \left\|h'\right\|_2^2 = 4A \cdot 4\pi^2 B = 16\pi^2 AB, \tag{21.284}$$

as desired.

■

Exercise 21.8. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous and integrable function such that

$$\lim_{|x|\to\infty} |f(x)|\,|x|^n = \lim_{|t|\to\infty} \left|\widehat{f}(t)\right| |t|^n = 0 \tag{21.285}$$

for any $n \in \mathbb{Z}_{\ge 0}$. Prove or disprove: f belongs to the Schwartz class $\mathcal{S}(\mathbb{R})$.

Solution. First we show that $\widehat{f} \in L^1$. In fact, we prove a stronger statement.

Claim 21.9: For any $n \ge 0$, $|t|^n \, abs\widehat{f} \in L^1$.

Proof. Since $\left|\widehat{f}(t)\right| |t|^{n+2} \to 0$, we can find some R such that $\left|\widehat{f}(t)\right| |t|^{n+2} \le 1$ for $|t| > R$. Since $\widehat{f}\,|t|^n$ is continuous, it's bounded on $[-R,R]$, say by M. Therefore,

$$\int_{\mathbb{R}} \left|\widehat{f}\right| |t|^n = \int_{[-R,R]} \left|\widehat{f}(t)\right| |t|^n + \int_{|t|>R} \left|\widehat{f}(t)\right| |t|^n \tag{21.286}$$

$$\le 2RM + \int_{|t|>R} \frac{1}{|t|^2} = 2RM + \frac{2}{R} < \infty. \tag{21.287}$$

Therefore, $|t|^n \, abs\widehat{f} \in L^1$. □

So $\widehat{f} \in L^1$. By Inversion formula, since $f, \widehat{f} \in L^1$,

$$f(x) = \int_{\mathbb{R}} \widehat{f}(t) e^{2\pi i x t} \, \mathrm{d}t, \tag{21.288}$$

for a.e. $x \in \mathbb{R}$. But since f is continuous, and the RHS is a continuous function of x, we have that the equality holds everywhere.

Claim 21.10: $f \in C^\infty(\mathbb{R})$.

Proof. We prove that we can differentiate under the integral sign in (21.288) using Claim 20.2. Consider

$$F(x,t) = \widehat{f}(t) e^{2\pi i x t}. \tag{21.289}$$

Then $F(x,-)$ is L^1, since its absolute value is just $\left|\widehat{f}(t)\right|$. Furthermore, $F(-,y)$ is smooth. Also,

$$\partial^k F(x,y) = (2\pi i t)^k \widehat{f}(t) e^{2\pi i x t}, \tag{21.290}$$

so we have

$$\left|\partial^k F(x,y)\right| = (2\pi)^k |t|^k \left|\widehat{f}(t)\right|, \tag{21.291}$$

which is integrable by Claim 21.9. Therefore, all the hypotheses of Claim 20.2 are satisfied, so we can differentiate under the integral sign. □

So f is smooth, with its k-th derivative being

$$f^{(k)}(x) = \int_{\mathbb{R}} (2\pi i t)^k \widehat{f}(t) e^{2\pi i x t} \, \mathrm{d}t, \tag{21.292}$$

which is bounded (in absolute value) by

$$\left|f^{(k)}(x)\right| \leq (2\pi)^k \int_{\mathbb{R}} |t|^k \left|\widehat{f}(t)\right| \, \mathrm{d}t =: M_k. \tag{21.293}$$

We claim that $f \in \mathcal{S}(\mathbb{R})$. We prove by induction on k that

$$\sup_{x \in \mathbb{R}} |x|^n \left|f^{(k)}(x)\right| < \infty, \quad \text{for all } n \geq 0. \tag{21.294}$$

The base case $k = 0$ is given in the hypothesis: $|x|^n |f(x)| \to 0$ as $|x| \to \infty$. So there exists some R such that $|x|^n |f(x)| \leq 1$ for $|x| > R$. On $[-R, R]$, the continuous function $|x|^n |f(x)|$ is bounded, by, say M,

$$\sup_{x \in \mathbb{R}} |x|^n |f(x)| \leq \max\{1, M\} < \infty. \tag{21.295}$$

So the base case is done! Suppose it holds for some $k \geq 0$. Then

$$C_{k,n} = \sup_{x \in \mathbb{R}} |x|^n \left|f^{(k)}(x)\right| < \infty. \tag{21.296}$$

Now, by Taylor's theorem,

$$f^{(k)}(x+h) = f^{(k)}(x) + hf^{(k+1)}(x) + \frac{h^2}{2} f^{(k+2)}(\xi), \tag{21.297}$$

for some ξ between x and $x+h$.

$$f^{(k+1)}(x) = \frac{f^{(k)}(x+h) - f^{(k)}(x)}{h} - \frac{h}{2} f^{(k+2)}(\xi). \tag{21.298}$$

Now, assume $|x| \geq 2$, and set $h = |x|^{-n} \in (0,1]$. Then

$$|x| \leq |x+h| + 1 \leq 2|x+h| \implies |x|^n \leq 2^n |x+h|^n. \tag{21.299}$$

Then from (21.298), we have

$$|x|^n \left|f^{(k+1)}(x)\right| \leq |x|^n \frac{\left|f^{(k)}(x+h)\right| + \left|f^{(k)}(x)\right|}{h} + \frac{h}{2}|x|^n \left|f^{(k+2)}(\xi)\right| \tag{21.300}$$

$$\leq |x|^{2n} \left|f^{(k)}(x+h)\right| + |x|^{2n} \left|f^{(k)}(x)\right| + \frac{1}{2} M_{k+2} \tag{21.301}$$

$$\leq 4^n |x+h|^n \left|f^{(k)}(x+h)\right| + |x|^{2n} \left|f^{(k)}(x)\right| + \frac{1}{2} M_{k+2} \tag{21.302}$$

$$\leq (4^n + 1) C_{k,2n} + \frac{1}{2} M_{k+2} < \infty. \tag{21.303}$$

$|x|^n \left|f^{(k+1)}(x)\right|$ is continuous, so it's bounded by M' on $[-2,2]$. Therefore,

$$\sup_{x \in \mathbb{R}} |x|^n \left|f^{(k+1)}(x)\right| \leq \max\left\{M', (4^n+1) C_{k,2n} + \frac{1}{2} M_{k+2}\right\} < \infty, \tag{21.304}$$

proving that $f \in \mathcal{S}$. ■

Exercise 21.9. In 1975, W. Beckner established the sharp form of the Hausdorff–Young inequality:

$$\left\|\widehat{f}\right\|_{L^{p'}} \leq C(p) \|f\|_{L^p}, \tag{21.305}$$

where $1 \leq p \leq 2$ and

$$C(p) = \left[\frac{p^{\frac{1}{p}}}{(p')^{\frac{1}{p'}}}\right]^{\frac{d}{2}} \tag{21.306}$$

Differentiating the sharp form of the Hausdorff–Young inequality (with respect to the parameter p) at $p = 2$, show that if $f \in \mathcal{S}$ with $\|f\|_{L^2} = 1$, then

$$-\int_{\mathbb{R}^d} (\log|f(x)|) |f(x)|^2 \, \mathrm{d}x - \int_{\mathbb{R}^d} \left(\log\left|\widehat{f}(\xi)\right|\right) \left|\widehat{f}(\xi)\right|^2 \mathrm{d}\xi \geq \frac{d}{2}(1 - \log 2). \tag{21.307}$$

Solution. Let $g : [1,2] \to \mathbb{R}$ be defined as

$$g(p) = C(p) \|f\|_{L^p} - \left\|\widehat{f}\right\|_{L^{p'}}. \tag{21.308}$$

Then $g \geq 0$, and $g(2) = 0$ by the isometry property. So we have $g(p) \geq g(2)$ for $p < 2$. So we have

$$g'(2) = \lim_{p \to 2^-} \frac{g(p) - g(2)}{p - 2} \leq 0. \tag{21.309}$$

Claim 21.11: For any $u \in \mathcal{S}$ with $\|u\|_2 = 1$, if we define $N(p) = \|u\|_p$, then

$$N'(2) = \frac{1}{2} \int_{\mathbb{R}^d} |u(x)|^2 \log |u(x)| \, \mathrm{d}x. \tag{21.310}$$

Proof. Using $N(p) = \left[\int_{\mathbb{R}^d} |u|^p\right]^{1/p}$, we have

$$\log N(p) = \frac{1}{p} \log \left[\int_{\mathbb{R}^d} |u|^p\right]. \tag{21.311}$$

Then by differentiating with respect to p, we get

$$\frac{N'(p)}{N(p)} = -\frac{1}{p^2} \log \left[\int_{\mathbb{R}^d} |u|^p\right] + \frac{1}{p} \frac{\frac{\mathrm{d}}{\mathrm{d}p} \int_{\mathbb{R}^d} |u|^p}{\int_{\mathbb{R}^d} |u|^p}. \tag{21.312}$$

Now, by putting $p = 2$, we have

$$N'(2) = \frac{1}{2} \left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} \int_{\mathbb{R}^d} |u|^p. \tag{21.313}$$

Now, for $p \in \left[\frac{3}{2}, \frac{5}{2}\right]$, one can bound $\frac{\mathrm{d}}{\mathrm{d}p}(|u|^p) = |u|^p \log |u|$ as follows: for $|u| \geq 1$,

$$|u|^p \log |u| \leq |u|^{p+1} \leq |u|^{\frac{7}{2}}. \tag{21.314}$$

For $|u| \leq 1$, we have

$$|u|^p \log |u| = |u|^{\frac{3}{2}} |u|^{p-\frac{3}{2}} \log |u| \leq C |u|^{\frac{3}{2}}, \tag{21.315}$$

since $\sup_{t \in (0,1]} |t^\varepsilon \log t| = \frac{1}{e\varepsilon}$. Therefore,

$$|u|^p \log |u| \leq u^{\frac{7}{2}} + C |u|^{\frac{3}{2}}. \tag{21.316}$$

Since $u \in \mathcal{S}$, it's in all L^p. Therefore, $|u|^p \log |u|$ is bounded by some integrable function. So we can differentiate under the integral sign, so

$$N'(2) = \frac{1}{2} \int_{\mathbb{R}^d} \left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} |u|^p . N'(2) = \frac{1}{2} \int_{\mathbb{R}^d} |u|^2 \log |u|. \tag{21.317}$$

□

So we have

$$\left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} \|f\|_p = \frac{1}{2}\int_{\mathbb{R}^d} |f|^2 \log|f| \,. \tag{21.318}$$

$$\left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} \left\|\widehat{f}\right\|_{p'} = \left[\left.\frac{\mathrm{d}}{\mathrm{d}p'}\right|_{p'=2} \left\|\widehat{f}\right\|_{p'}\right] \left[\left.\frac{\mathrm{d}p'}{\mathrm{d}p}\right|_{p=2}\right] = -\frac{1}{2}\int_{\mathbb{R}^d} \left|\widehat{f}\right|^2 \log\left|\widehat{f}\right| , \tag{21.319}$$

as $p' = \frac{p}{p-1}$, so $\frac{\mathrm{d}p'}{\mathrm{d}p} = -\frac{1}{(p-1)^2}$. Now,

$$\log C(p) = \frac{d}{2}\left(\frac{\log p}{p} - \frac{\log p'}{p'}\right). \tag{21.320}$$

Differentiating with respect to p, we have

$$\frac{C'(p)}{C(p)} = \frac{d}{2}\left[\frac{1-\log p}{p^2} - \frac{1-\log p'}{p'^{\,2}}\frac{\mathrm{d}p'}{\mathrm{d}p}\right] = \frac{d}{2}\left[\frac{1-\log p}{p^2} + \frac{1-\log p'}{p'^{\,2}(p-1)^2}\right]. \tag{21.321}$$

Putting $p = 2$, we have

$$C'(2) = \frac{d}{2}\,\frac{1-\log 2}{2} = \frac{d}{4}\,(1-\log 2)\,. \tag{21.322}$$

So we have

$$g'(2) = C'(2)\,\|f\|_2 + C(2)\,\left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} \|f\|_p - \left.\frac{\mathrm{d}}{\mathrm{d}p}\right|_{p=2} \left\|\widehat{f}\right\|_{p'} \tag{21.323}$$

$$= \frac{d}{4}\,(1-\log 2) + \frac{1}{2}\int_{\mathbb{R}^d} |f|^2 \log|f| + \frac{1}{2}\int_{\mathbb{R}^d} \left|\widehat{f}\right|^2 \log\left|\widehat{f}\right| . \tag{21.324}$$

Since $g'(2) \le 0$, after rearranging, we have

$$\frac{d}{2}\,(1-\log 2) \le -\int_{\mathbb{R}^d} |f|^2 \log|f| - \int_{\mathbb{R}^d} \left|\widehat{f}\right|^2 \log\left|\widehat{f}\right| . \tag{21.325}$$

■

References

[1] W. Beckner, *Inequalities in Fourier Analysis*, Ann. Math. 102 (1975), 159–182.

[2] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, John Wiley and Sons Inc., 1999.

[3] L. Grafakos, *Classical and Modern Fourier Analysis*, Pearson Education Inc., 2004.

[4] E. H. Lieb and M. Loss, *Analysis*, Graduate Studies in Mathematics, Volume 14, 2nd edition, American Mathematical Society, 2001.

[5] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1966.

[6] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[7] E. M. Stein, *Harmonic Analysis*, Princeton University Press, 1993.

[8] E. M. Stein and G. Weiss, *Introduction to Fourier Analysis on Euclidean spaces*, Princeton University Press, 1973.

[9] E. M. Stein and R. Shakarchi, *Real Analysis: Measure theory, Integration and Hilbert Spaces*, Princeton Lecture Series in Analysis III, Princeton University Press, 2005.

[10] R. L. Wheeden and A. Zygmund, *Measure and Integral*, Monographs and textbooks in pure and applied mathematics, Marcel Dekker, Inc., New York, 1977.

[11] A. Zygmund, *Trigonometric Series, Vol II*, Cambridge University Press, 1959.