INTRODUCTION TO MEASURE THEORY

MAYUKH MUKHERJEE

Contents

1. Review of basics 1
1.1. Why pass beyond Riemann integration? 1
1.2. From content to Lebesgue measure 2
1.3. Integration on measure spaces 6
2. The Caratheodory construction 11
2.1. Metric spaces 13
3. Product measures 14
4. Examples of measures 16
4.1. Basic constructions 16
4.2. Measures selected by symmetry and structure 21
4.3. Classical geometric measures: Besicovitch, De Giorgi, and Federer 31
4.4. Three cautions and counterexamples 39
5. Signed and complex measures 39
6. L^p spaces 41
6.1. A few remarks around the heat equation 52
7. Appendix: Bare bones of distribution theory 54
7.1. Fourier transform and tempered distributions 54
8. Radon measures and the Riesz-Kakutani represention theorem 58
8.1. A few remarks and immediate consequences 61
9. Sobolev spaces and almost everywhere differentiability 62
10. Hausdorff measures 69
10.1. Isoperimetric inequalities 76
10.2. Co-area formula 76
References 87

These notes are mainly aimed at providing a PhD level "introduction" to Measure theory. We will begin with a quick review of basic measure theory that you are already aware of. For more details, please have a look at [LL01, Roy88, Tay06].

1. Review of basics

1.1. **Why pass beyond Riemann integration?** Lebesgue theory is designed to remedy two closely related limitations of Riemann integration. The Riemann integral is organised around finite partitions, and consequently does not interact well with general pointwise limits. The corresponding notion of the size of a set, namely Jordan content, is likewise organised around finite coverings. Since limiting arguments throughout analysis naturally involve countable operations, both limitations point towards the same enlargement of the theory. Throughout, I denotes a non-degenerate bounded closed interval.

First consider convergence of functions. A uniformly bounded, pointwise increasing sequence of Riemann-integrable functions need not have a Riemann-integrable limit. Indeed, enumerate

$$I \cap \mathbb{Q} = \{x_1, x_2, \dots\}$$

and set

$$\varphi_k = \chi_{\{x_1,\dots,x_k\}}.$$

Each φ_k is Riemann integrable and $0 \le \varphi_k \le 1$, whereas

$$\varphi_k \nearrow \chi_{I\cap\mathbb{Q}}.$$

The limiting function is not Riemann integrable, since both $I \cap \mathbb{Q}$ and $I \setminus \mathbb{Q}$ are dense in I. Thus the class of Riemann-integrable functions is not closed under bounded monotone pointwise limits. Lebesgue integration enlarges the class of admissible functions precisely so that such limits are governed by the Monotone Convergence Theorem.

The corresponding difficulty for sets is already visible in the notion of content.

Definition 1.1. *Let $S \subseteq I$. Its upper and lower Jordan contents are*

$$\mathrm{cont}^+(S) := \mathcal{U}(\chi_S), \qquad \mathrm{cont}^-(S) := \mathcal{L}(\chi_S),$$

where $\mathcal{U}$ and $\mathcal{L}$ denote the upper and lower Riemann integrals, respectively. We say that S has Jordan content if $\mathrm{cont}^+(S) = \mathrm{cont}^-(S)$. In that case their common value is denoted by $\mathrm{cont}(S)$.

Equivalently, S has Jordan content precisely when χ_S is Riemann integrable, in which case

$$\mathrm{cont}(S) = \int_I \chi_S.$$

Exercise 1.2. *Prove that*

$$\mathrm{cont}^+(S) = \inf\left\{\sum_{k=1}^{N} |J_k| : S \subseteq \bigcup_{k=1}^{N} J_k, \quad J_1, \dots, J_N \subseteq I \text{ are intervals}\right\}.$$

The essential restriction here is that only finite coverings are allowed. For example,

$$\mathrm{cont}^+(I \cap \mathbb{Q}) = |I|,$$

because every non-degenerate subinterval of I contains a rational number. This calculation is correct, but it is incompatible with the notion of size that we seek: a countable set should have measure zero, since singletons have measure zero and a measure should be countably additive.

The two difficulties therefore have the same source. Finite partitions and finite coverings are not sufficiently stable under countable limiting operations. Lebesgue theory resolves this by beginning with countable coverings, thereby producing an outer measure, and then identifying the sets on which that outer measure behaves additively.

1.2. **From content to Lebesgue measure.** Motivated by the preceding discussion, we replace the finite coverings in the definition of upper content by countable coverings. For $S \subseteq I$, define the Lebesgue outer measure of S by

$$m^*(S) = \inf\left\{\sum_{k=1}^{\infty} |J_k| : S \subseteq \bigcup_{k=1}^{\infty} J_k, \quad J_k \subseteq I \text{ are intervals}\right\}.$$

Thus Lebesgue outer measure is the countable analogue of upper Jordan content. The passage from finite to countable coverings is exactly what allows countable sets, including $I \cap \mathbb{Q}$, to have outer measure zero.

Exercise 1.3. *Check that the above definition remains unaffected if J_k are open/closed/half-open.*

We prove the useful property of countable subadditivity for outer measure.

Proposition 1.4. *Let $S_j, j \geq 0$ be a countable family of subsets of I, then*

$$m^*\left(\bigcup_j S_j\right) \leq \sum m^*(S_j). \tag{1}$$

Proof. For each j, take a countable cover J_{j_k} of S_j such that $\sum_k |J_{j_k}| \leq m^*(S_j) + \frac{\varepsilon}{2^j}$. Now, clearly $\{J_{j_k}\}$ acts as a cover for $\cup_j S_j$. Then,

$$m^*\left(\bigcup_j S_j\right) \leq \sum_{j,k} |J_{j_k}| \leq \sum m^*(S_j) + \varepsilon.$$

□

Somewhat expectedly, we now define the concept of inner measure.

Definition 1.5. *Let $S \subseteq I$. Then the inner measure of S is defined by*

$$m_*(S) = |I| - m^*(I \setminus S). \tag{2}$$

Observe that $m_*(S) \leq m^*(S)$ (why?).

Definition 1.6. *A set $S \subseteq I$ is called measurable if $m^*(S) = m_*(S)$.*

Exercise 1.7. *Let $S \subseteq I$. Prove the following:*

$$m^*(S) = \inf\{m^*(U) : S \subseteq U, U \text{ open }\}, \tag{3}$$

and,

$$m_*(S) = \sup\{m^*(K) : K \subseteq S, K \text{ compact}\}. \tag{4}$$

Similar to Proposition 1.4, we have countable superadditivity of inner measure:

Proposition 1.8. *Let $\{S_k\}_{k\geq 1}$ be a countable family of pairwise disjoint subsets of I. Then*

$$m_*\left(\bigcup_{k\geq 1} S_k\right) \geq \sum_{k\geq 1} m_*(S_k).$$

Proof. Fix $\varepsilon > 0$. By the characterisation of inner measure in (4), for each $k \geq 1$ there exists a compact set $K_k \subseteq S_k$ such that

$$m(K_k) \geq m_*(S_k) - 2^{-k}\varepsilon.$$

Recall that compact sets are measurable, since (4) gives $m_*(K_k) = m^*(K_k)$.

For $N \geq 1$, set

$$F_N := \bigcup_{k=1}^{N} K_k.$$

This is compact and therefore measurable. We claim that

$$m(F_N) = \sum_{k=1}^{N} m(K_k).$$

Indeed, after discarding any empty K_k, the finitely many pairwise disjoint compact sets $K_1, \ldots, K_N$ have positive mutual separation. Given any interval covering of F_N, subdivide its intervals, without changing their total length, so that every resulting interval has length smaller than this separation.

Each resulting interval can then meet at most one of the sets K_k. Grouping the intervals according to the set K_k that they meet shows that

$$m^*(F_N) \geq \sum_{k=1}^{N} m^*(K_k).$$

The reverse inequality follows from finite subadditivity. Since F_N and the K_k are measurable, the claimed equality follows.

Since $F_N \subseteq \bigcup_{k\geq 1} S_k$, monotonicity of inner measure gives

$$m_*\left(\bigcup_{k\geq 1} S_k\right) \geq m_*(F_N) = \sum_{k=1}^{N} m(K_k) \geq \sum_{k=1}^{N} m_*(S_k) - \varepsilon \sum_{k=1}^{N} 2^{-k}.$$

Letting $N \to \infty$ and using $\sum_{k\geq 1} 2^{-k} = 1$, we obtain

$$m_*\left(\bigcup_{k\geq 1} S_k\right) \geq \sum_{k\geq 1} m_*(S_k) - \varepsilon.$$

Finally, let $\varepsilon \searrow 0$. □

Putting Propositions 1.4 and 1.8, we get the following important

Theorem 1.9. *If $S_j \subseteq I$ be a countable family of mutually disjoint measurable subsets of I, then $S = \bigcup_{j\geq 1} S_j$ is measurable, and*

$$m\left(\bigcup_j S_j\right) = \sum_j m(S_j). \tag{5}$$

Proof. It is clear that

$$m^*(S) \leq \sum_{j\geq 1} m^*(S_j) = \sum_{j\geq 1} m_*(S_j) \leq m_*(S).$$

□

The following is a useful fact which we leave as an exercise.

Exercise 1.10. *Prove that a set $S \subseteq I$ is measurable if and only if, for each $\delta > 0$, there exist a compact K and an open U such that $K \subseteq S \subseteq U$ and $m(U \setminus K) < \delta$.*

Here is another useful exercise:

Exercise 1.11. *If $S_1, S_2 \subseteq I$ are measurable, then so are $S_1 \cup S_2, S_1 \cap S_2, S_1 \setminus S_2$.*

We prove the countable analogue of the last exercise:

Theorem 1.12. *If $S_j \subseteq I$ is a countable collection of measurable sets, then $\bigcup_{j\geq 1} S_j, \bigcap_{j\geq 1} S_j$ are also measurable.*

Proof. It suffices only to prove the claim for $U := \bigcup_{j\geq 1} S_j$, as $\bigcap_{j\geq 1} S_j = I \setminus \bigcup_{j\geq 1}(I \setminus S_j)$. Let $S'_j := \bigcup_{1\leq k\leq j} S_k$, and $T_j := S'_j \setminus S'_{j-1}$. Since the T_j are mutually disjoint, $U = \bigcup_j T_j$ is measurable. □

Now we come to the following important

Definition 1.13. *A non-empty family $\mathcal{F}$ of subsets of a set X is called a σ-algebra if it is closed under formation of countable unions and intersections, and also complements.*

We have already proved that the family of measurable subsets of I is a σ-algebra. The *smallest* σ-algebra containing all the closed sets is called the algebra of Borel sets.

Exercise 1.14. *Prove that the above mentioned concept of Borel sets is well-defined.*

Proposition 1.15 (Carathéodory's criterion)**.** *A set $S \subseteq I$ is measurable if and only if, for all $Y \subseteq I$,*

$$m^*(Y) = m^*(Y \cap S) + m^*(Y \setminus S). \tag{6}$$

Proof. One side is clear by taking $Y = I$. So we assume that S is measurable. Observe that $m^*(Y) \leq m^*(Y \cap S) + m^*(Y \setminus S)$ always holds. Given $\varepsilon > 0$, choose an open set U containing Y such that $m(U) \leq m^*(Y) + \varepsilon$. Since $U \cap S$ and $U \setminus S$ are measurable, we have that

$$m(U) = m(U \cap S) + m(U \setminus S) \geq m^*(Y \cap S) + m^*(Y \setminus S).$$

Hence, $m^*(Y \cap S) + m^*(Y \setminus S) \leq m^*(Y) + \varepsilon$. □

Now, we define the Lebesgue measure on the real line $\mathbb{R}$. Partition $\mathbb{R} = \bigcup_k I_k$, so that two I_ks can intersect at at most one point. We say $S \subseteq \mathbb{R}$ is measurable if and only if $S \cap I_k$ is measurable for all k. If this holds, then we set

$$m(S) = \sum_k m(S \cap I_k). \tag{7}$$

Exercise 1.16. *Show that the above definition of measurability is independent of the partition $\{I_k\}$.*

1.2.1. *Non-measurable sets.* The Lebesgue σ-algebra cannot contain every subset of $\mathbb{T}$ while preserving both countable additivity and translation invariance. To see this, endow

$$\mathbb{T} = \mathbb{R}/\mathbb{Z}$$

with normalized Lebesgue measure, so that $m(\mathbb{T}) = 1$. The countable group $\mathbb{Q}/\mathbb{Z}$ acts on $\mathbb{T}$ by translations. Equivalently, define

$$x \sim y \quad \iff \quad x - y \in \mathbb{Q}/\mathbb{Z}.$$

Using the axiom of choice, select a set $S \subseteq \mathbb{T}$ containing exactly one representative from each equivalence class. For $\alpha \in \mathbb{Q}/\mathbb{Z}$, set

$$S_\alpha = S + \alpha.$$

Then

$$\mathbb{T} = \bigsqcup_{\alpha \in \mathbb{Q}/\mathbb{Z}} S_\alpha.$$

Indeed, every point of $\mathbb{T}$ belongs to some translate of S, while two distinct translates cannot intersect because S contains only one representative from each equivalence class.

Suppose that S were measurable. By translation invariance, every S_α would then be measurable and

$$m(S_\alpha) = m(S).$$

Since $\mathbb{Q}/\mathbb{Z}$ is countably infinite, countable additivity would give

$$1 = m(\mathbb{T}) = \sum_{\alpha \in \mathbb{Q}/\mathbb{Z}} m(S_\alpha) = \sum_{\alpha \in \mathbb{Q}/\mathbb{Z}} m(S).$$

This is impossible: the last sum is 0 if $m(S) = 0$ and is infinite if $m(S) > 0$. Hence S is not measurable. Moreover, none of its translates S_α is measurable, since measurability of one translate would imply measurability of S after translation by $-\alpha$.

Remark 1.17. *A much more appealing way of showing the existence of non-measurable sets is via the Banach-Tarski paradox*[1]*.*

[1]This will be a presentation topic for the midsems!

1.3. **Integration on measure spaces.** Start with a set X and a σ-algebra $\mathcal{F}$ of subsets of X. A measure μ on $(X, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ satisfying

$$\mu(\emptyset) = 0, \tag{8}$$

and

$$S_j \in \mathcal{F} \text{ disjoint, countable } \implies \mu\left(\bigcup_j S_j\right) = \sum_j \mu(S_j). \tag{9}$$

We call $(X, \mathcal{F}, \mu)$ a measure space. It is immediate that μ has monotonicity and countable subadditivity. Also,

Exercise 1.18. *Prove that if $S_j \in \mathcal{F}$, $S_1 \subseteq S_2 \subseteq S_3 \subseteq \ldots$, then $\mu(S_j) \nearrow \mu\left(\bigcup S_j\right)$.*

Definition 1.19. *A function $f : X \to \mathbb{R}$ (respectively, $[-\infty, \infty]$) is called $\mathcal{F}$-measurable if $f^{-1}(S) \in \mathcal{F}$ for all Borel sets $S \subseteq \mathbb{R}$ (respectively, $[-\infty, \infty]$).*

The following are left as exercises:

Exercise 1.20 (Algebra of measurable functions)**.** *Prove the following:*

(1) *If $f_j : X \to \mathbb{R}$ are measurable, then so are $f_1 + f_2$ and $f_1 f_2$.*
(2) *More generally, if f_j are measurable, for $j = 1, \ldots, n$, and $\Phi : \mathbb{R}^n \to \mathbb{R}$ is continuous, then $g : X \to \mathbb{R}$, defined by*
$$g(x) = \Phi(f_1(x), \ldots, f_n(x))$$
is measurable.
(3) *If f_j are measurable for $j \in \mathbb{N}$, then $\sup_j f_j(x), \inf_j f_j(x), \limsup_j f_j(x), \liminf f_j(x)$ are all measurable functions (from X to $[-\infty, \infty]$).*

If $(X, \mathcal{F}, \mu)$ is a measure space and $S \in \mathcal{F}$, then χ_S is $\mathcal{F}$-measurable, and so is any finite linear combination of such functions. Any such finite linear combination

$$\varphi(x) = \sum_{j=1}^{N} a_j \chi_{S_j}(x), a_j \in \mathbb{R}, S_j \in \mathcal{F}$$

is called a *simple function.*

Exercise 1.21. *Prove that any simple function can be written as a linear combination of characteristic functions of disjoint sets. (Hint: for each subset $\sigma \subseteq \{1, \ldots, N\}$, define*

$$\tilde{S}_\sigma = \left(\bigcap_{j \in \sigma} S_j\right) \cap \left(\bigcap_{j \notin \sigma} X \setminus S_j\right).$$

Then check that $\varphi(x) = \sum_\sigma \tilde{a}_\sigma \chi_{\tilde{S}_\sigma}(x)$, where $\tilde{a}_\sigma = \sum_{j \in \sigma} a_j$.)

Denote by $\mathfrak{S}^+(X)$ the class of simple functions that are non-negative on X. By the above exercise, they can be written in the form $\sum_{j=1}^N a_j \chi_{S_j}(x)$, where $a_j \geq 0$.

Definition 1.22. *Given $\varphi \in \mathfrak{S}^+(X)$, define the integral*

$$\int_X \varphi \, d\mu = \sum_{j=1}^{N} a_j \mu(S_j). \tag{10}$$

Remark 1.23. *By convention, we take $a_j \mu(S_j) = 0$ if $a_j = 0$, regardless of whether $\mu(S_j)$ is finite or not. Also, the above definition satisfies the obvious properties of additivity and monotonicity, as well as the expected interaction with scalar multiplication. We skip the formalities.*

Now, let $\mathcal{M}^+(X)$ denote the set of all measurable functions $f : X \to [0, \infty]$. Now we define the integral of a non-negative measurable function.

Definition 1.24. *For $f \in \mathcal{M}^+(X)$, we define the integral*

$$\int_X f \, d\mu = \sup \left\{ \int_X \varphi \, d\mu : 0 \leq \varphi \leq f, \varphi \in \mathfrak{S}^+(X) \right\}. \tag{11}$$

Now, we start proving the Monotone Convergence Theorem, which was advertised in the introduction, and is one of the central results of Lebesgue theory.

Theorem 1.25 (Monotone Convergence Theorem)**.** *If $f_j \in \mathcal{M}^+(X)$ and $f_j \nearrow$, then $f_j \nearrow f \in \mathcal{M}^+(X)$ and*

$$\int f \, d\mu = \lim_{j \to \infty} \int f_j \, d\mu. \tag{12}$$

Proof. Clearly, $\lim_{j \nearrow \infty} f_j(x)$ exists (maybe $+\infty$-valued) and the limit is measurable via Exercise 1.20. By monotonicity, we have $\int f_j \leq \int f$, which implies that $\lim_j \int f_j \leq \int f$. For the reverse inequality, take $\alpha \in (0, 1)$, and let $f \geq \varphi \in \mathfrak{S}^+(X)$. Let $E_n := \{x \in X : f_n(x) \geq \alpha\varphi\}$. Then, E_n is an increasing sequence of measurable sets whose union is X, and

$$\lim \int f_j \geq \int_{E_j} f_j \geq \alpha \int_{E_j} \varphi. \tag{13}$$

We leave the following as an exercise (see below):

$$\lim \int_{E_j} \varphi = \int \varphi. \tag{14}$$

This gives that $\lim_j \int f_j \geq \alpha \int \varphi$ for all $\alpha \in (0, 1)$. Letting $\alpha \to 1$, we see that $\lim_j \int f_j \geq \int \varphi$ for all $\varphi \leq f$. This proves the claim. □

Exercise 1.26. *Prove (14) above. (Hint: Observe that if $\varphi \in \mathfrak{S}^+(X)$, then $\int_A \varphi$ is a measure on $\mathcal{F}$. Also use Exercise 1.18).*

Remark 1.27. *The following is a "default" approximation method by step functions. Given $f \in \mathcal{M}^+(X)$, define the sequence φ_j by*

$$\varphi_j(x) = \begin{cases} k/2^j, \text{ on } \{x \in X : k2^{-j} \leq f(x) < (k+1)2^{-j}, k < j.2^j\} \\ j \text{ on } \{x \in X : j \leq f(x)\}. \end{cases}$$

One can check that $\varphi_j \nearrow f$.

The next is another landmark result in Lebesgue theory: the Monotone Convergence Theorem immediately gives the following lower-semicontinuity property of the integral.

Proposition 1.28 (Fatou's lemma)**.** *If $f_j \in \mathcal{M}^+(X)$, then*

$$\int_X \liminf_{j \to \infty} f_j \, d\mu \leq \liminf_{j \to \infty} \int_X f_j \, d\mu.$$

Proof. For each $k \geq 1$, define

$$g_k := \inf_{j \geq k} f_j.$$

Then g_k is measurable and non-negative, and

$$g_k \nearrow \liminf_{j \to \infty} f_j$$

pointwise. Moreover, for every $j \geq k$ we have $g_k \leq f_j$, and hence

$$\int_X g_k \, d\mu \leq \inf_{j \geq k} \int_X f_j \, d\mu.$$

Applying the Monotone Convergence Theorem to $\{g_k\}$ gives

$$\begin{aligned} \int_X \liminf_{j \to \infty} f_j \, d\mu &= \lim_{k \to \infty} \int_X g_k \, d\mu \\ &\leq \lim_{k \to \infty} \inf_{j \geq k} \int_X f_j \, d\mu \\ &= \liminf_{j \to \infty} \int_X f_j \, d\mu. \end{aligned}$$

□

Remark 1.29. *Suppose f_j converges to f a.e. Then, Fatou's lemma claims that*

$$\int f \leq \liminf \int f_j.$$

Even in this case the inequality can be strict. Consider the sequence of functions $f_j(x) = 1/j$ for $|x| \leq j$ and $f_j = 0$ otherwise. Intuitively, this is an example where the "mass" of the functions escapes out to infinity (a phenomenon well-known to people familiar with precompactness arguments in the calculus of variations). It turns out (by a famous work of Brezis and Lieb) that this is the only essential problem for which Fatou's lemma can be non-sharp. For more details, consult Theorem 1.9 of [LL01].

Definition 1.30. *For a measurable function $f : X \to \mathbb{R}$, write $f = f^+ - f^-$. Clearly, $f^\pm \in \mathcal{M}^+(X)$, so the integral of f is defined as*

$$\int f = \int f^+ - \int f^-. \tag{15}$$

If both $\int f^+$ and $\int f^-$ are finite, then f is said to be integrable.

If $f : X \in \mathbb{C}$, write $f = u + iv$ for real-valued u, v and define

$$\int f = \int u + i \int v. \tag{16}$$

Now, we prove the Dominated Convergence Theorem.

Theorem 1.31 (Dominated Convergence Theorem)**.** *Let f_n be integrable for $n \in \mathbb{N}$. Suppose $f : X \to \mathbb{R}$ and $f_n \to f$ for all $x \in X$, and suppose there is an integrable function g such that $|f_n| \leq g$ for all n. Then, f is integrable, and $\int f = \lim \int f_n$.*

Proof. By a previous exercise, f is measurable, and $|f|$ is integrable, as g is. Since $g + f_n \geq 0$, by Fatou's lemma, it is clear that

$$\begin{aligned} \int g + \int f &= \int g + f \\ &\leq \liminf \int (g + f_n) \\ &= \int g + \liminf \int f_n, \end{aligned}$$

and hence,

$$\int f \leq \liminf \int f_n.$$

Replacing f_n by $-f_n$, we get analogously,

$$\int f \geq \limsup \int f_n.$$

□

The next theorem establishes a very important bridge between pointwise and uniform convergence:

Theorem 1.32 (Egorov's theorem)**.** *Let $(X, \mathcal{F}, \mu)$ be a finite measure space. Let f_j be measurable and $f_j(x) \to f(x)$ for all $x \in X$. Then, given any $\delta > 0$, there exists a $B \in \mathcal{F}$ such that $\mu(B) \leq \delta$ and $f_j \to f$ uniformly on $X \setminus B$.*

Proof. For $\varepsilon > 0, n \in \mathbb{N}$, define

$$F_{n,\varepsilon} = \{x \in X : \text{ for some } j \geq n, |f_j(x) - f(x)| \geq \varepsilon\}. \tag{17}$$

Note that each $F_{n,\varepsilon}$ is measurable. One can easily check that $\bigcap_{n\geq 1} F_{n,\varepsilon} = \emptyset$, so $\mu(F_{n,\varepsilon}) \to 0$ as $n \to \infty$.

In particular, given any $\delta > 0$, set $\varepsilon = 2^{-k}$, whence one can find $n(k)$ such that $\mu\left(F_{n(k),2^{-k}}\right) \leq 2^{-k}\delta$. If $B = \bigcup_{k\geq 1} F_{n(k),2^{-k}}$, we leave it to the reader to check that B satisfies the claim of the theorem. □

We recall the following insight from [Roy88], which talks about Littlewood's three principles. Roughly, the three principles are the following:

(1) Every measurable set is nearly a finite union of intervals
(2) Every convergent sequence of measurable functions is nearly uniformly convergent.
(3) Every measurable function is nearly continuous

Now, a liberal interpretation of the first principle is given by Exercise 1.10 above. Also, an illustration of the second principle is given by Theorem 1.32 above. Now we provide an illustration of the third principle. Various variants of the following are known as Lusin's theorem.

Exercise 1.33 (Lusin's theorem)**.** *Prove the following:*

(1) *Let f be an integrable function on I. Show that there exist $f_j \in C(I)$ such that $f_j \to f$ almost everywhere. (Hint: Use approximation by simple functions first).*
(2) *Given $\delta > 0$, show that there is a Lebesgue measurable $S \subseteq I$ such that $m(S) < \delta$ and $f|_{I\setminus S}$ is continuous. (Hint: apply Egorov's theorem).*

Remark 1.34. *Lusin's theorem could be stated more generally on metric measure spaces. You can think about this extension.*

Remark 1.35. *We mention a point which is important for applications. In all the convergence theorems established above, the hypothesis that a sequence of measurable functions f_j converges for all $x \in X$ can be weakened to the hypothesis that f_j converges almost everywhere. The proof proceeds by considering $X_1 \subseteq X$ such that $\mu(X \setminus X_1) = 0$, and restricting attention to the measure space $(X_1, \mathcal{F}_1, \mu_1)$, where $\mathcal{F}_1 = \{S \cap X_1 : S \in \mathcal{F}\}$ and μ_1 is μ restricted to $\mathcal{F}_1$.*

We now come full circle by relating the Lebesgue integral to the Riemann integral.

Theorem 1.36 (Lebesgue's criterion for Riemann integrability)**.** *Let $\varphi : I \to \mathbb{R}$ be bounded. Then φ is Riemann integrable if and only if its set of discontinuities has Lebesgue measure zero. In that case, φ is Lebesgue integrable and its Riemann and Lebesgue integrals agree.*

Proof. Let $\mathcal{P}_0 = \{I\}$, and let $\mathcal{P}_{j+1}$ be obtained by bisecting each interval in $\mathcal{P}_j$. For $x \in I$, let $C_j(x)$ be the union of all intervals in $\mathcal{P}_j$ containing x, and define

$$u_j(x) := \sup_{y \in C_j(x)} \varphi(y), \qquad \ell_j(x) := \inf_{y \in C_j(x)} \varphi(y).$$

This convention removes the ambiguity at common endpoints of adjacent intervals. The functions u_j and ℓ_j are measurable, and

$$u_j \searrow u \geq \varphi, \qquad \ell_j \nearrow \ell \leq \varphi$$

pointwise.

Except at the finitely many endpoints of $\mathcal{P}_j$, the functions u_j and ℓ_j are the usual upper and lower Darboux step functions. Consequently,

$$\int_I u_j \, dm = \mathcal{U}(\varphi, \mathcal{P}_j), \qquad \int_I \ell_j \, dm = \mathcal{L}(\varphi, \mathcal{P}_j).$$

Since the mesh of $\mathcal{P}_j$ tends to zero,

$$\mathcal{U}(\varphi, \mathcal{P}_j) \searrow \mathcal{U}(\varphi), \qquad \mathcal{L}(\varphi, \mathcal{P}_j) \nearrow \mathcal{L}(\varphi).$$

Indeed, if $\mathcal{P}$ is any fixed partition, compare $\mathcal{P}_j$ with the common refinement $\mathcal{P} \cup \mathcal{P}_j$. Only those intervals of $\mathcal{P}_j$ meeting an endpoint of $\mathcal{P}$ can contribute an error, and their total length tends to zero.

All the functions u_j and ℓ_j are bounded by a common integrable constant. Hence the Dominated Convergence Theorem gives

$$\mathcal{U}(\varphi) = \int_I u \, dm, \qquad \mathcal{L}(\varphi) = \int_I \ell \, dm.$$

Since $u \geq \ell$, it follows that

$$\begin{aligned} \varphi \text{ is Riemann integrable} &\iff \mathcal{U}(\varphi) = \mathcal{L}(\varphi) \\ &\iff \int_I (u - \ell) \, dm = 0 \\ &\iff u = \ell \quad \text{almost everywhere.} \end{aligned}$$

Let E be the collection of endpoints of all the partitions $\mathcal{P}_j$. This is a countable set and hence has measure zero. If $x \notin E$, there is a unique interval $J_j(x) \in \mathcal{P}_j$ containing x, and x lies in its interior. Moreover,

$$u(x) - \ell(x) = \lim_{j \to \infty} \left(\sup_{y \in J_j(x)} \varphi(y) - \inf_{y \in J_j(x)} \varphi(y) \right).$$

If φ is continuous at x, the expression on the right tends to zero because the intervals $J_j(x)$ shrink to x. Conversely, if it tends to zero, then for every $\varepsilon > 0$ some $J_j(x)$ has oscillation smaller than ε. Since x lies in the interior of $J_j(x)$, this implies that φ is continuous at x. Therefore, for every $x \notin E$,

$$u(x) = \ell(x) \quad \iff \quad \varphi \text{ is continuous at } x.$$

Since E has measure zero, $u = \ell$ almost everywhere if and only if the set of discontinuities of φ has measure zero.

Finally, whenever these equivalent conditions hold,

$$\ell \leq \varphi \leq u \qquad \text{and} \qquad u = \ell \quad \text{almost everywhere.}$$

Thus $\varphi = u = \ell$ almost everywhere, so φ is Lebesgue integrable, and

$$\int_I \varphi \, dm = \int_I u \, dm = \mathcal{U}(\varphi) = \mathcal{L}(\varphi),$$

which is precisely its Riemann integral. □

2. The Caratheodory construction

Recall that an outer measure on a set X is a function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ satisfying

$$\mu^*(\emptyset) = 0, \tag{18}$$

$$A \subseteq B \implies \mu^*(A) \leq \mu^*(B) \tag{19}$$

$$A_j \subseteq X \text{ countable} \implies \mu^*\left(\bigcup_j A_j\right) \leq \sum_j \mu^*(A_j). \tag{20}$$

Let $\mathcal{E}$ be a family of subsets of X, such that $\emptyset \in \mathcal{E}$, and $X = \bigcup_{j\geq 1} X_j$ for some countable collection $X_j \in \mathcal{E}$. Suppose there is a function $\rho : \mathcal{E} \to [0, \infty]$ such that $\rho(\emptyset) = 0$. Then for any $S \subseteq X$, set

$$\mu^*(S) = \inf\left\{\sum_{j\geq 1} \rho(E_j) : E_j \in \mathcal{E}, S \subseteq \bigcup_{j\geq 1} E_j\right\}. \tag{21}$$

The following is parallel to the proofs for Lebesgue measure we have seen before:

Proposition 2.1. *(21) gives an outer measure on X.*

Now, suppose we define a set $A \subseteq X$ to be measurable if

$$\mu^*(Y) = \mu^*(Y \cap A) + \mu^*(Y \setminus A), \quad \forall Y \subseteq X, \tag{22}$$

and denote the set of μ^*-measurable subsets of X by $\mathcal{M}$. Then,

Theorem 2.2 (Caratheodory's theorem)**.** *$\mathcal{M}$, as defined above, is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a complete[2] measure.*

The proof is straightforward.

Proof. Clearly, $\emptyset \in \mathcal{M}$. Also, if $A \in \mathcal{M}$, then for all $Y \subseteq X$, $Y \cap A^c = Y \setminus A$ and $Y \setminus A^c = Y \cap A$, so $\mathcal{M}$ is closed under complements.

Now, suppose that $A_j \in \mathcal{M}$. We want to show that (22) holds with $A = A_1 \cup A_2$. Since μ^* is subadditive, it suffices to show that

$$\mu^*(Y) \geq \mu^*(Y \cap (A_1 \cup A_2)) + \mu^*(Y \cap (A_1 \cup A_2)^c)$$

for all $Y \subseteq X$. Note that $A_1 \cup A_2 = A_1 \cup (A_2 \cap A_1^c)$. Then,

$$\begin{aligned}\mu^*(Y \cap (A_1 \cup A_2)) + \mu^*(Y \cap (A_1 \cup A_2)^c) &\leq \mu^*(Y \cap A_1) + \mu^*(Y \cap A_2 \cap A_1^c) + \mu^*(Y \cap A_1^c \cap A_2^c) \\ &= \mu^*(Y \cap A_1) + \mu^*(Y \cap A_1^c) = \mu^*(Y).\end{aligned}$$

This shows that $\mathcal{M}$ is an algebra[3].

We now check additivity when $A_1, A_2 \in \mathcal{M}$ are disjoint. Setting $Y = A_1 \cup A_2$, $A = A_1$ in (22), we get

$$\mu^*(A_1 \cup A_2) = \mu^*(A_1) + \mu^*(A_2), A_j \in \mathcal{M} \text{ disjoint.}$$

Inductively, for $A_j \in \mathcal{M}$ disjoint,

$$\mu^*\left(\bigcup_{j=1}^N A_j\right) = \sum_{j=1}^N \mu^*(A_j),$$

[2]Every subset of a measure zero set is also measurable

[3]Recall that this means that $\mathcal{M}$ is closed under complements and finite unions.

and taking $N \nearrow \infty$ and using monotonicity of μ^*, yield

$$\mu^*\left(\bigcup_{j\geq 1} A_j\right) \geq \sum_{j\geq 1} \mu^*(A_j),$$

and since the reverse inequality follows from the definition of outer measure, we have countable additivity.

It remains to show that countable union of disjoint measurable sets is measurable. Let $B_n = \bigcup_{j\leq n} A_j$. For each n, one has

$$\begin{aligned}\mu^*(Y \cap B_n) &= \mu^*(Y \cap B_n \cap A_n) + \mu^*(Y \cap B_n \cap A_n^c) \\ &= \mu^*(Y \cap A_n) + \mu^*(Y \cap B_{n-1}),\end{aligned}$$

for any $Y \subseteq X$. Inductively, we obtain

$$\mu^*(Y \cap B_n) = \sum_{j=1}^{n} \mu^*(Y \cap A_j).$$

We know that $\mathcal{M}$ is an algebra, so $B_n \in \mathcal{M}$, and hence using the above, we have

$$\begin{aligned}\mu^*(Y) &= \mu^*(Y \cap B_n) + \mu^*(Y \cap B_n^c) \\ &\geq \sum_{j=1}^{n} \mu^*(Y \cap A_j) + \mu^*(Y \cap A^c).\end{aligned}$$

Taking $n \to \infty$, we have

$$\begin{aligned}\mu^*(Y) &\geq \sum_{j\geq 1} \mu^*(Y \cap A_j) + \mu^*(Y \cap A^c) \\ &\geq \mu^*\left(\bigcup_{j\geq 1} (Y \cap A_j)\right) + \mu^*(Y \cap A^c) \\ &= \mu^*(Y \cap A) + \mu^*(Y \cap A^c) \geq \mu^*(Y).\end{aligned}$$

Thus $A \in \mathcal{M}$. □

We now relate $\mu^*(S)$ to $\rho(S)$ when $S \in \mathcal{E}$. Let $\mathcal{A}$ be an algebra[4]. A function $\mu_0 : \mathcal{A} \to [0, \infty]$ is called a pre-measure if it satisfies

$$\mu_0(\emptyset) = 0, \tag{23}$$

and countable additivity. A pre-measure induces an outer measure via (21).

Proposition 2.3. *If μ_0 is a premeasure inducing the outer measure μ^*, then*

$$S \in \mathcal{A} \implies \mu^*(S) = \mu_0(S) \tag{24}$$

and every set in $\mathcal{A}$ is μ^*-measurable.

Proof. Firstly, $\mu^*(S) \leq \mu_0(S)$, since S covers itself. Suppose $S \subseteq \bigcup_{j\geq 1} A_j$, where $A_j \in \mathcal{A}$. We see that $B_n = S \cap \left(A_n \setminus \bigcup_{j<n} A_j\right)$ is a disjoint family inside $\mathcal{A}$. Now, countable additivity gives us

$$\mu_0(S) = \sum_{j\geq 1} \mu_0(B_j) \leq \sum_{j\geq 1} \mu_0(A_j), \tag{25}$$

which proves that $\mu_0(S) \leq \mu^*(S)$, which proves (24).

[4]nonempty collection of subsets of X, closed under finite unions and complements.

To prove that each $A \in \mathcal{A}$, choose an arbitrary $Y \subseteq X$. For each $\varepsilon > 0$, there is a sequence $\{B_j, j \geq 1\} \subseteq \mathcal{A}$ such that $Y \subseteq \bigcup_{j\geq 1} B_j$ and $\sum \mu_0(B_j) \leq \mu^*(Y) + \varepsilon$. Since μ_0 is additive on $\mathcal{A}$,

$$\begin{aligned}\mu^*(Y) + \varepsilon &\geq \sum_{j\geq 1} \mu_0(B_j \cap A) + \sum_{j\geq 1} \mu_0(B_j \cap A^c) \\ &\geq \mu^*(Y \cap A) + \mu^*(Y \cap A^c).\end{aligned}$$

Letting $\varepsilon \searrow 0$, we prove our claim. □

Putting together Theorem 2.2 and Proposition 2.3, we obtain the following:

Theorem 2.4. *Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be an algebra, μ_0 a premeasure on $\mathcal{A}$, and $\mathcal{M} = \sigma(\mathcal{A})$ be the σ-algebra generated by $\mathcal{A}$. Then there exists a measure μ on $\mathcal{M}$ whose restriction to $\mathcal{A}$ is μ_0.*

We leave the uniqueness issue as an exercise:

Exercise 2.5. *Let $\mathcal{A}$ be an algebra generating the σ-algebra $\mathcal{M}$. Let μ_0 be a premeasure on $\mathcal{A}$ and μ be a measure on $\mathcal{M}$ given by Theorem 2.4 extending μ_0. If ν is another measure on $\mathcal{M}$ which agrees with μ_0 on $\mathcal{A}$, then for all $S \in \mathcal{M}$, $\nu(S) \leq \mu(S)$ and equality holds when $\mu(S) < \infty$. Furthermore, if there is a countable family A_j such that*

$$X = \bigcup_{j\geq 1} A_j, \quad A_j \in \mathcal{M}, \quad \mu(A_j) < \infty,$$

then $\mu(S) = \nu(S)$ for all $S \in \mathcal{M}$.

2.1. **Metric spaces.** Let X be a metric space, with distance function $d(x, y)$. An outer measure μ^* on X is called a *metric outer measure* provided

$$\rho(S_1, S_2) = \inf\{d(x_1, x_2) : x_j \in S_j\} > 0 \implies \mu^*(S_1 \cup S_2) = \mu^*(S_1) + \mu^*(S_2). \tag{26}$$

Exercise 2.6. *Prove that the Lebesgue outer measure is a metric outer measure on $I \subseteq \mathbb{R}$.*

Here is a generalisation of the above exercise due to Caratheodory, which is a very efficient method of making sure that Borel sets are measurable:

Proposition 2.7. *If μ^* is a metric outer measure on a metric space X, then every closed subset of X is μ^*-measurable.*

Proof. Suppose $F \subseteq X$ is closed and $\mu^*(Y) < \infty$. It suffices to prove that

$$\mu^*(Y) \geq \mu^*(Y \cap F) + \mu^*(Y \setminus F). \tag{27}$$

Consider $B_n := \{x \in Y \setminus F : \rho(x, F) \geq 1/n\}$. Clearly, $B_n \nearrow Y \setminus F$. Also, $\rho(B_n, F) \geq 1/n$, implying that $\mu^*(Y \cap F) + \mu^*(B_n) = \mu^*((Y \cap F) \cup B_n) \leq \mu^*(Y)$. So it suffices to show that

$$\mu^*(B_n) \to \mu^*(Y \setminus F). \tag{28}$$

Let $C_n = B_{n+1} \setminus B_n$. Note that $|j - k| \geq 2 \implies \rho(C_j, C_k) > 0$. Hence, one obtains inductively that

$$\sum_{j=1}^{N} \mu^*(C_{2j}) = \mu^*\left(\bigcup_{j=1}^{N} C_{2j}\right) \leq \mu^*(Y), \tag{29}$$

$$\sum_{j=1}^{N} \mu^*(C_{2j+1}) = \mu^*\left(\bigcup_{j=1}^{N} C_{2j+1}\right) \leq \mu^*(Y), \tag{30}$$

and consequently $\sum_{j\geq 1}\mu^*(C_j) < \infty$[5]. Countable subadditivity implies

$$\mu^*(Y\setminus F) \leq \mu^*(B_n) + \sum_{j\geq n}\mu^*(C_j),$$

so as $n\to\infty$, the latter tend to zero, and we obtain

$$\mu^*(Y\setminus F) \leq \liminf_{n\to\infty}\mu^*(B_n) \leq \limsup_{n\to\infty}\mu^*(B_n) \leq \mu^*(Y\setminus F).$$

□

3. Product measures

We start with a few preliminaries.

Definition 3.1. *A monotone class $\mathfrak{M}$ is a collection of sets with two properties:*

(1) *if $A_i\in\mathfrak{M}, i\in\mathbb{N}$, and if $A_1\subseteq A_2\subseteq\ldots$, then $\bigcup_i A_i\in\mathfrak{M}$,*
(2) *if $B_j\in\mathfrak{M}, i\in\mathbb{N}$ and if $B_1\supseteq B_2\supseteq\ldots$, then $\bigcap_j B_j\in\mathfrak{M}$.*

We state the following algebraic fact without proof, and refer the reader to Theorem 1.3 of [LL01], for example.

Theorem 3.2 (Monotone Class Theorem)**.** *Let X be a set and let $\mathcal{A}$ be an algebra of subsets of X such that $X,\emptyset\in\mathcal{A}$. Then there exists a smallest monotone class $\mathcal{S}$ containing $\mathcal{A}$, which is also the smallest σ-algebra containing $\mathcal{A}$.*

If $(X,\mathcal{M},\mu),(Y,\mathcal{N},\nu)$ are measure spaces, then we define a *rectangle* as a set of the form $A\times B\subseteq X\times Y$, where $A\in\mathcal{M}, B\in\mathcal{N}$. One can check that the collection of finite disjoint unions of rectangles is an algebra, which we denote by $\mathcal{M}\boxtimes\mathcal{N}$. The σ-algebra generated by the latter is denoted by $\mathcal{M}\otimes\mathcal{N}$. We define the set function $\pi:\mathcal{M}\boxtimes\mathcal{N}\to[0,\infty]$ by

$$\pi\left(\bigcup_{j=1}^{N}(A_j\times B_j)\right) = \sum_{j=1}^{N}\mu(A_j)\nu(B_j), \tag{31}$$

when the rectangles $R_j = A_j\times B_j$ are disjoint.

Exercise 3.3. *Prove that π is a premeasure on $\mathcal{M}\boxtimes\mathcal{N}$.*

Now we define the important concept of outer measure on $X\times Y$ in the following way: for $S\subseteq X\times Y$, define

$$\pi^*(S) = \inf\left\{\sum_{k\geq 1}\mu(A_k)\nu(B_k) : A_k\in\mathcal{M}, B_k\in\mathcal{N}, S\subseteq\bigcup_{k\geq 1}(A_k\times B_k)\right\}. \tag{32}$$

As you can realise, the above definition is parallel to (21).

We leave the following as an exercise:

Exercise 3.4. *Prove the following:*

(1) *The restriction of π^* to $\mathcal{M}\otimes\mathcal{N}$ is a measure, which we call the product measure $\mu\times\nu$.*
(2) *$\mu\times\nu$ is the unique measure on $\mathcal{M}\times\mathcal{N}$ with the following projection property*

$$(\mu\times\nu)(A\times B) = \mu(A)\nu(B), A\in\mathcal{M}, B\in\mathcal{N}. \tag{33}$$

(Hint: Use Theorem 2.4 and Exercise 2.5 above).

[5]These steps were required. Continuity from below does not rescue the argument as B_n are not automatically measurable.

Our main aim in this section is to relate the iterated integral $\int \left(\int f(x,y)\, d\nu(y) \right)\, d\mu(x)$ with the double integral $\int f(x,y) d(\mu \times \nu)$. For a set $S \subseteq X \times Y$, define

$$S_x = \{y \in Y : (x,y) \in S\}, S_y = \{x \in X : (x,y) \in S\}.$$

Lemma 3.5. *If $S \in \mathcal{M} \otimes \mathcal{N}$, then $S_x \in \mathcal{N}, S_y \in \mathcal{M}$. More generally, if f is $\mathcal{M} \otimes \mathcal{N}$-measurable, then $f(x,.)$ (respectively $f(.,y)$) is $\mathcal{N}$-measurable (respectively $\mathcal{M}$-measurable) for all $x \in X$ (respectively, for all $y \in Y$).*

Proof. Let $\mathcal{F} = \{S \subseteq X \times Y : S_x \in \mathcal{N} \quad \forall x \in X, S_y \in \mathcal{M} \quad \forall y \in Y\}$. Then one can easily check that $\mathcal{F}$ contains the rectangles and is closed under countable unions and intersections. This means that $\mathcal{F}$ is a σ-algebra containing $\mathcal{M} \otimes \mathcal{N}$. Since $f(x,.)^{-1}(S) = (f^{-1}(S))_x$, the rest is clear. □

Now we prove the characteristic function version of the Fubini-Tonelli theorem. Recall that a measure μ on X is called σ-finite if X can be expressed as a countable union of sets of finite measure.

Proposition 3.6. *If μ, ν are σ-finite, $S \in \mathcal{M} \otimes \mathcal{N}$, then*

$$\varphi(x) = \nu(S_x), \psi(y) = \mu(S_y) \text{ are measurable }, \tag{34}$$

and

$$(\mu \times \nu)(S) = \int \varphi(x)\, d\mu = \int \psi(y)\, d\nu, \tag{35}$$

Proof. Assume that μ, ν are finite measures. Let $\mathcal{C}$ denote the collection of sets $S \subseteq X \times Y$ which satisfy (34), (35). We wish to prove that $\mathcal{C} = \mathcal{M} \otimes \mathcal{N}$. Clearly, $\mathcal{C}$ contains the finite disjoint unions of rectangles. If we could show that $\mathcal{C}$ is a monotone class, then the Monotone Class Theorem would prove our contention.

Let $S_j \in \mathcal{C}$, $S_j \nearrow S$. Then their projections are also monotone, and $\psi_j(y) = \mu(S_{j_y}) \nearrow \psi(y) = \mu(S_y)$, and the MCT implies that

$$\int \mu(S_y)\, d\nu = \lim \int \psi_j(y)\, d\nu = \lim (\mu \times \nu)(S_j) = (\mu \times \nu)(S). \tag{36}$$

Similarly for S_x.

On the other hand, if $S_j \in \mathcal{C}$ and $S_j \searrow S$, then $\psi_j(y) \searrow \psi(y)$, provided μ is finite. Since ν is finite, we can apply DCT (because $\psi_j(y) = \mu(S_{j_y}) \leq \mu(X)$) and get (36). □

Exercise 3.7. *Extend the above proof to the case of σ-finite measures.*

Theorem 3.8 (Tonelli's Theorem)**.** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces. If $f \in \mathcal{M}^+(X \times Y)$, then*

$$g(x) = \int f_x\, d\nu \in \mathcal{M}^+(X), h(y) = \int f_y\, d\mu \in \mathcal{M}^+(Y) \tag{37}$$

and

$$\int f\, d(\mu \times \nu) = \int g(x)\, d\mu = \int h(y)\, d\nu. \tag{38}$$

Proof. We note that from Proposition 3.6, the theorem is clear for simple functions f_j. Now, take a sequence of simple functions $f_j \nearrow f$, whence $f_{j_x} \nearrow f_x$, and $f_{j_y} \nearrow f_y$. Now MCT gives the claim. □

Theorem 3.9 (Fubini's Theorem)**.** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are σ-finite measure spaces. If $f \in L^1(X \times Y, \mu \times \nu)$, then*

$$f_y \in L^1(X, \mu) v - a.e., f_x \in L^1(Y, \nu) \mu - a.e. \tag{39}$$

Furthermore,

$$g(x) = \int f_x \, d\nu \in L^1(X, \mu), h(y) = \int f_y \, d\mu \in L^1(Y, \nu) \tag{40}$$

and

$$\int f \, d(\mu \times \nu) = \int g(x) \, d\mu = \int h(y) \, d\nu. \tag{41}$$

Proof. The proof is immediate from Theorem 3.8 by splitting $f = f^+ - f^-$. □

4. Examples of measures

We pause the abstract development to see what measures look like in different parts of mathematics. The examples also warn against excessive Euclidean intuition. A measure need not have a density, be non-atomic, be complete, or be σ-finite.

4.1. **Basic constructions.**

4.1.1. *Point masses, counting measure, and empirical measures.* For $x \in X$, the *Dirac measure* at x is defined on $\mathcal{P}(X)$ by

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

Thus integration against δ_x is evaluation at x:

$$\int_X f \, d\delta_x = f(x) \tag{42}$$

for every non-negative measurable function f, and also for every δ_x-integrable function.

Exercise 4.1. *Prove* (42).

More generally, if $x_j \in X$ are distinct and $a_j \geq 0$, then

$$\mu = \sum_{j \geq 1} a_j \delta_{x_j}, \qquad \mu(A) = \sum_{\{j : x_j \in A\}} a_j,$$

is a measure, and

$$\int_X f \, d\mu = \sum_{j \geq 1} a_j f(x_j)$$

for non-negative measurable f, and also for integrable f. Measure-theoretic integration therefore contains summation as a special case.

The *counting measure* on an arbitrary set X is

$$\#(A) = \begin{cases} \text{the number of elements of } A, & A \text{ is finite}, \\ +\infty, & A \text{ is infinite}. \end{cases}$$

When L^p spaces are introduced later, counting measure will give the familiar sequence spaces $\ell^p(X)$ for $X = \mathbb{N}$. If X is uncountable, however, counting measure is not σ-finite: every set of finite counting measure is finite, and a countable union of finite sets cannot cover X. Thus σ-finiteness is a genuine additional property of a measure, rather than a consequence of the measure axioms.

If $x_1, \ldots, x_N$ are observations in a measurable space, then

$$\mu_N := \frac{1}{N} \sum_{j=1}^{N} \delta_{x_j}$$

is their *empirical measure*. For any measurable f,

$$\int f \, d\mu_N = \frac{1}{N} \sum_{j=1}^{N} f(x_j).$$

Thus the passage from finite samples to limiting distributions is literally the passage from averages of Dirac masses to limits of measures. The same construction occurs in statistics, dynamical systems, and the study of particle systems.

Remark 4.2. *Dirac's "delta function" arose in the development of quantum mechanics as an idealized unit mass concentrated at a point. As a measure, it is a perfectly ordinary object. It also defines a distribution by $\varphi \mapsto \varphi(x)$, a point of view to which we return in the appendix on distribution theory.*

4.1.2. *Transporting a measure: pushforwards and laws.* Many important measures are obtained not by prescribing their values directly, but by transporting a measure that is already understood.

Definition 4.3. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $(Y, \mathcal{N})$ be a measurable space, and let $T : X \to Y$ be measurable. The* pushforward *of μ under T is the measure $T_{\#}\mu$ on $(Y, \mathcal{N})$ defined by*

$$(T_{\#}\mu)(B) := \mu(T^{-1}(B)), \qquad B \in \mathcal{N}.$$

The notation is meant to suggest that T pushes mass on X forward to Y. Notice that inverse images appear in the definition. They preserve complements, countable unions, and disjointness, precisely the set operations needed to transfer the σ-algebra structure and countable additivity through T. Direct images preserve unions, but generally preserve neither complements nor disjointness.

Proposition 4.4 (Integration against a pushforward)**.** *If $f : Y \to [0, \infty]$ is measurable, then*

$$\int_Y f(y) \, d(T_{\#}\mu)(y) = \int_X f(T(x)) \, d\mu(x).$$

The same identity holds for every integrable $f : Y \to \mathbb{R}$.

Proof. The identity follows immediately for characteristic functions from the definition of $T_{\#}\mu$. It follows for non-negative simple functions by linearity, and then for arbitrary non-negative measurable functions by monotone convergence. Applying the result to the positive and negative parts proves the integrable case. □

If Z is a real-valued random variable (measurable function) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then its *law* is the probability measure

$$\mathrm{Law}(Z) := Z_{\#}\mathbb{P}.$$

The preceding proposition becomes

$$\mathbb{E}[f(Z)] = \int_{\mathbb{R}} f(t) \, d\,\mathrm{Law}(Z)(t).$$

This is why one may study a random variable through its distribution without retaining the original probability space. Pushforwards also underlie changes of variables, surface measures, and the spectral measures considered below.

We shall use two standard comparisons between measures on the same measurable space. We write $\nu \ll \mu$ if every μ-null set is also ν-null, and $\nu \perp \mu$ if there is a measurable set E such that $\nu(X \setminus E) = 0$ and $\mu(E) = 0$. These are called absolute continuity and mutual singularity, respectively. Their systematic study comes later with the Radon–Nikodym and Lebesgue decomposition theorems.

4.1.3. *Coin tossing, Lebesgue measure, and Cantor measure.* Let

$$\Omega = \{0,1\}^{\mathbb{N}}.$$

For $\varepsilon_1, \dots, \varepsilon_n \in \{0,1\}$, define the cylinder

$$[\varepsilon_1 \cdots \varepsilon_n] := \{\omega \in \Omega : \omega_1 = \varepsilon_1, \dots, \omega_n = \varepsilon_n\}.$$

There is a unique probability measure $\mathbb{P}$ on the σ-algebra generated by the cylinders such that

$$\mathbb{P}([\varepsilon_1 \cdots \varepsilon_n]) = 2^{-n}.$$

Exercise 4.5. *The above prescription defines a premeasure on the algebra of finite unions of cylinders, and the Carathéodory extension theorem applies. Prove this.*

Consider first the binary coding map

$$b(\omega) := \sum_{j=1}^{\infty} \frac{\omega_j}{2^j}.$$

Then

$$b_{\#}\mathbb{P} = m|_{[0,1]}.$$

To see the idea, fixing the first n digits specifies a dyadic interval of length 2^{-n}, and the corresponding cylinder also has probability 2^{-n}. The two possible binary expansions of a dyadic rational create only a null ambiguity. Let $\nu = b_{\#}\mathbb{P}$. For every half-open dyadic interval

$$I_{n,k} = \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right),$$

the set $b^{-1}(I_{n,k})$ agrees, up to finitely many $\mathbb{P}$-null sequences, with the corresponding cylinder of length n. Hence

$$\nu(I_{n,k}) = 2^{-n} = m(I_{n,k}).$$

Since the dyadic intervals generate $\mathcal{B}([0,1])$ and ν and m are finite measures, the uniqueness theorem for measures implies $\nu = m|_{[0,1]}$.

Now change only the coding map, setting

$$\kappa(\omega) := \sum_{j=1}^{\infty} \frac{2\omega_j}{3^j}, \qquad \nu_C := \kappa_{\#}\mathbb{P}.$$

The image of κ is the middle-thirds Cantor set C. Every basic interval remaining after n stages of the Cantor construction has ν_C-measure 2^{-n}. More precisely, for $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{0,1\}^n$, the corresponding basic interval is

$$I_\varepsilon = \left[\sum_{j=1}^{n} \frac{2\varepsilon_j}{3^j}, \sum_{j=1}^{n} \frac{2\varepsilon_j}{3^j} + \frac{1}{3^n}\right].$$

Since

$$\kappa^{-1}(I_\varepsilon) = [\varepsilon_1 \cdots \varepsilon_n],$$

we have

$$\nu_C(I_\varepsilon) = \mathbb{P}([\varepsilon_1 \cdots \varepsilon_n]) = 2^{-n}.$$

It follows that

$$\nu_C(C) = 1, \qquad m(C) = 0,$$

so ν_C is singular with respect to Lebesgue measure. Since $\kappa(\{0,1\}^{\mathbb{N}}) = C$ and $\nu_C = \kappa_{\#}\mathbb{P}$, we have

$$\nu_C(C) = \mathbb{P}\big(\kappa^{-1}(C)\big) = \mathbb{P}(\{0,1\}^{\mathbb{N}}) = 1.$$

Thus ν_C is concentrated on the Cantor set, even though C has Lebesgue measure zero. On the other hand, ν_C has no atoms: if $x \in C$, then for every n it lies in a basic interval of ν_C-measure 2^{-n}, and hence

$$0 \le \nu_C(\{x\}) \le 2^{-n} \qquad \text{for every } n.$$

Thus $\nu_C(\{x\}) = 0$.

This is the *Cantor measure*. It is simultaneously singular and non-atomic, and is therefore a basic counterexample to the tempting claim that every non-atomic measure on $\mathbb{R}$ has a density. If

$$S_0(x) = \frac{x}{3}, \qquad S_1(x) = \frac{x+2}{3},$$

then it also satisfies the self-similarity relation

$$\nu_C = \frac{1}{2}(S_0)_{\#}\nu_C + \frac{1}{2}(S_1)_{\#}\nu_C.$$

This identity is the prototype for self-similar measures in fractal geometry.

The same coin-toss measure has therefore produced both Lebesgue measure and a singular measure, merely by changing the map through which it is viewed. This is one reason pushforwards are so useful.

Remark 4.6. *Kolmogorov's 1933 axiomatization made probability measures on infinite product spaces a foundation for stochastic processes. In the present example, the coordinate maps $\omega \mapsto \omega_j$ are independent random variables, while the coding maps turn the entire sequence into a random point of $[0,1]$ or of the Cantor set.*

A discrete path-space precursor. The same cylinder construction gives the law of a random walk on a graph. Let $G = (V, E)$ be a locally finite d-regular graph, fix $o \in V$, and set

$$\Omega_o = \big\{(v_0, v_1, \dots) : v_0 = o \text{ and } v_{j+1} \sim v_j \text{ for every } j \ge 0\big\}.$$

For an admissible finite path $\gamma = (o, v_1, \dots, v_n)$, let

$$[\gamma] = \{\omega \in \Omega_o : \omega_j = v_j \text{ for } 0 \le j \le n\}$$

and prescribe

$$\mathbb{P}_o([\gamma]) = d^{-n}.$$

The required consistency is transparent:

$$[\gamma] = \bigsqcup_{w \sim v_n} [\gamma, w], \qquad d^{-n} = \sum_{w \sim v_n} d^{-(n+1)}.$$

Consequently, the prescription is well-defined and finitely additive on the algebra of finite unions of path cylinders.

It is also a premeasure. Indeed, since G is locally finite and the initial vertex is fixed, the ball $B(o, n)$ is finite for every n. Therefore Ω_o is a closed subset of the compact product

$$\prod_{n \ge 0} B(o, n),$$

and every path cylinder is both open and closed. If an algebra set is a disjoint countable union of algebra sets, compactness reduces this union to a finite one. Countable additivity therefore reduces to finite additivity. The Carathéodory extension theorem now gives a unique probability measure $\mathbb{P}_o$ on the σ-algebra generated by the path cylinders.

More generally, if $p(x, y)$ are transition probabilities, then one sets

$$\mathbb{P}_o([o, v_1, \dots, v_n]) = \prod_{j=0}^{n-1} p(v_j, v_{j+1}).$$

The consistency identity is then precisely $\sum_y p(x,y) = 1$. In the d-regular case there is also a direct pushforward description. Label the neighbours of every vertex by $1, \dots, d$ and follow successively the edges specified by an element of $\{1, \dots, d\}^{\mathbb{N}}$. This map pushes the uniform product measure on $\{1, \dots, d\}^{\mathbb{N}}$ forward to $\mathbb{P}_o$.

This is usually called the *random-walk path measure*, rather than Wiener measure. It is nevertheless a useful discrete precursor to the construction of Wiener measure. There the transition probabilities are replaced by the heat kernel, consistency follows from the Chapman–Kolmogorov identity, and one must additionally prove that the resulting measure is concentrated on continuous paths.

4.1.4. *Lebesgue–Stieltjes measures.* The preceding examples can be placed in a common one-dimensional framework.

Theorem 4.7 (Lebesgue–Stieltjes measure)**.** *Let $F : \mathbb{R} \to \mathbb{R}$ be non-decreasing and right-continuous. There is a unique Borel measure μ_F, finite on bounded intervals, such that*

$$\mu_F((a,b]) = F(b) - F(a), \qquad a < b.$$

Conversely, if μ is a finite Borel measure on $\mathbb{R}$, then its distribution function

$$F_\mu(x) := \mu((-\infty, x])$$

is non-decreasing and right-continuous, and $\mu = \mu_{F_\mu}$.

The theorem is another direct application of the Carathéodory extension theorem: first define the desired premeasure on finite disjoint unions of half-open intervals.

Several familiar measures result from very different choices of F:

(1) If $F(x) = x$, then μ_F is Lebesgue measure.
(2) If $F = \chi_{[a,\infty)}$, then $\mu_F = \delta_a$.
(3) If $g \geq 0$ is locally integrable and
$$F(x) = \int_0^x g(t)\, dt,$$
with the usual signed interpretation when $x < 0$, then $d\mu_F = g\, dm$.
(4)
(5) Let
$$F_C(x) := \nu_C((-\infty, x]).$$
This is the *Cantor–Lebesgue function*, or simply the *Cantor function*. It is continuous and non-decreasing, and is constant on every connected component of $[0,1] \setminus C$. Moreover,
$$F_C(x) = 0 \quad \text{for } x < 0, \qquad F_C(x) = 1 \quad \text{for } x > 1.$$
Since
$$F_C(b) - F_C(a) = \nu_C((a,b]),$$
the associated Lebesgue–Stieltjes measure is
$$\mu_{F_C} = \nu_C.$$

Moreover,

$$\mu_F(\{x\}) = F(x) - F(x-), \qquad F(x-) := \lim_{t \nearrow x} F(t).$$

Thus jumps of F are precisely atoms of the measure. A continuous distribution function gives a non-atomic measure, but the Cantor function shows that continuity does not imply the existence of a density. These examples display the three basic behaviours of a finite measure on the line: atomic, absolutely continuous, and singular continuous.

Remark 4.8. *Stieltjes introduced integration with respect to an increasing function in his 1894–1895 memoir on continued fractions and the moment problem: from the moments of a mass distribution, one seeks to recover the distribution itself. The modern Lebesgue–Stieltjes measure is the measure-theoretic form of that construction. In probability theory it says, in particular, that every distribution function determines a probability measure.*

4.2. **Measures selected by symmetry and structure.**

4.2.1. *Haar measure.* Lebesgue measure is not merely a convenient assignment of lengths and volumes. It is, up to normalization, the only reasonable measure on $\mathbb{R}^n$ that is invariant under translations. Haar's theorem extends this principle from Euclidean space to groups.

Theorem 4.9 (Haar)**.** *Let G be a locally compact Hausdorff topological group. There exists a non-zero Borel measure m_G, finite on compact sets and regular, such that*

$$m_G(gA) = m_G(A)$$

for every Borel set $A \subseteq G$ and every $g \in G$. Any two such left-invariant measures differ by multiplication by a positive constant. If G is compact, there is therefore a unique such measure normalized by $m_G(G) = 1$.

The existence proof is substantial and will not be given here. The theorem turns the informal phrase "choose a group element uniformly" into a mathematically meaningful statement whenever the group is compact. Some examples are:

(1) On a discrete group, Haar measure is counting measure, up to a constant. On a finite group, normalized Haar measure assigns mass $1/|G|$ to each element.
(2) On $(\mathbb{R}^n, +)$, Haar measure is a constant multiple of Lebesgue measure.
(3) On the circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, normalized Haar measure is normalized arc length. It is the measure used implicitly in the Vitali example on the circle.
(4) On the multiplicative group $(0, \infty)$, Haar measure is

$$\frac{dx}{x}.$$

Indeed, the substitution $x = ay$ shows that

$$\int_{aA} \frac{dx}{x} = \int_A \frac{dy}{y}.$$

Thus the measure natural for scaling is different from the measure natural for translation.
(5) The additive group of the p-adic numbers has a Haar measure that may be normalized by $m_p(\mathbb{Z}_p) = 1$, where $\mathbb{Z}_p$ is the compact-open subgroup of p-adic integers. For $k \geq 0$, the group $p^k\mathbb{Z}_p$ has index p^k in $\mathbb{Z}_p$, so

$$m_p(p^k\mathbb{Z}_p) = p^{-k}.$$

This measure is fundamental in number theory and p-adic analysis.

Haar introduced this measure in 1933 in order to integrate compatibly with the translation structure of continuous groups; Weil subsequently placed the theory in its modern generality. Haar measure is the foundation of convolution, harmonic analysis, and Fourier analysis on groups. We will encounter it again when interpreting n-dimensional Hausdorff measure as Lebesgue measure in Section 10.

4.2.2. *Riemannian volume and Liouville measure.* Geometric structure also selects natural measures. Let (M, g) be an n-dimensional Riemannian manifold. In local coordinates $x^1, \dots, x^n$, write $g = (g_{ij})$. The *Riemannian volume measure* is locally

$$dV_g = \sqrt{\det(g_{ij})}\, dx^1 \cdots dx^n.$$

The factor $\sqrt{\det(g_{ij})}$ is exactly what compensates for a change of coordinates, so these local expressions define a global measure. Every Riemannian isometry preserves dV_g. This is the natural measure in geometric analysis: the Laplace–Beltrami operator, heat kernel, Sobolev inequalities, and geometric flows are all integrated against it. To verify the coordinate invariance directly, let $x = (x^1, \dots, x^n)$ and $y = (y^1, \dots, y^n)$ be two coordinate systems on the same open set, and put

$$J = \frac{\partial x}{\partial y}.$$

The matrices of the metric in the two coordinate systems are related by

$$G_y = J^T G_x J.$$

It follows that

$$\det G_y = (\det J)^2 \det G_x, \qquad \sqrt{\det G_y} = |\det J| \sqrt{\det G_x}.$$

Since the change-of-variables formula gives

$$dx^1 \cdots dx^n = |\det J|\, dy^1 \cdots dy^n,$$

we obtain

$$\sqrt{\det G_y}\, dy^1 \cdots dy^n = \sqrt{\det G_x}\, dx^1 \cdots dx^n.$$

Hence the local expressions agree on overlapping coordinate charts and therefore define a global measure.

There is a related measure in Hamiltonian mechanics. If (M^{2n}, ω) is a symplectic manifold, the volume form

$$\frac{\omega^n}{n!}$$

defines the *Liouville measure*. In canonical coordinates (q, p) it is simply

$$dq_1 \cdots dq_n\, dp_1 \cdots dp_n.$$

Hamiltonian flows preserve this measure. This is Liouville's theorem: an ensemble of states moves through phase space like an incompressible fluid. It is one of the historical origins of measure-preserving dynamics and statistical mechanics, and it supplies an important class of examples for the later section on ergodic theorems.

4.2.3. *Gaussian measure: high-dimensional spherical shadows.* The standard Gaussian measure on $\mathbb{R}^n$ is

$$d\gamma_n(x) = (2\pi)^{-n/2} e^{-|x|^2/2}\, dx.$$

Its coordinates are independent standard Gaussians, and it is invariant under every orthogonal transformation. More generally, if $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $A_{\#}\gamma_n$ is Gaussian with covariance AA^T. If A is not surjective, this measure is supported on the proper subspace $\operatorname{Ran} A$ and is therefore singular with respect to m-dimensional Lebesgue measure.

Gaussian measure also arises as the low-dimensional shadow of high-dimensional spherical measure. Let $Z_1, Z_2, \dots$ be independent standard Gaussian random variables and set

$$\Theta^{(N)} = \frac{\sqrt{N}\,(Z_1, \dots, Z_N)}{(Z_1^2 + \cdots + Z_N^2)^{1/2}}.$$

Rotational invariance implies that $\Theta^{(N)}$ is uniformly distributed on the sphere $S^{N-1}(\sqrt{N})$. The law of large numbers gives

$$\frac{Z_1^2 + \cdots + Z_N^2}{N} \longrightarrow 1 \qquad \text{almost surely.}$$

Consequently, for every fixed k,

$$(\Theta_1^{(N)}, \dots, \Theta_k^{(N)}) \longrightarrow (Z_1, \dots, Z_k)$$

in distribution. Equivalently, if σ_N denotes normalized surface measure on $S^{N-1}(\sqrt{N})$ and $\pi_k : \mathbb{R}^N \to \mathbb{R}^k$ is coordinate projection, then

$$(\pi_k)_{\#}\sigma_N \rightharpoonup \gamma_k.$$

Thus a fixed-dimensional observer sees an increasingly high-dimensional sphere as Gaussian.

There is a complementary concentration phenomenon: for every $\varepsilon > 0$,

$$\gamma_N\left(\left\{x : (1-\varepsilon)\sqrt{N} \le |x| \le (1+\varepsilon)\sqrt{N}\right\}\right) \longrightarrow 1.$$

Although the Gaussian density is largest at the origin, almost all of its mass in high dimension lies near the sphere of radius $\sqrt{N}$.

For high enough dimensional spheres, another interesting phenomenon is that any small enough neighborhood of any great circle (equator) contains most of the surface area. This can be seen by integrating the polar volume form $\sin^{n-1} r dr d\theta$ and considering high enough n.

Infinite dimensions introduce a genuine obstruction. An infinite-dimensional Banach space is not locally compact and hence admits no analogue of Lebesgue measure that is both locally finite and translation invariant. Even the naive infinite-dimensional standard Gaussian fails on ℓ^2: the product measure ${\gamma_1}^{\times\mathbb{N}}$ exists on $\mathbb{R}^{\mathbb{N}}$, but

$$\frac{1}{N}\sum_{j=1}^{N} x_j^2 \longrightarrow 1 \qquad \text{almost surely,}$$

and therefore

$${\gamma_1}^{\times\mathbb{N}}(\ell^2) = 0.$$

More generally, a Gaussian Borel probability on a Hilbert space must have trace-class covariance, so identity covariance is impossible in infinite dimensions. Gaussian measures do nevertheless exist on suitable larger spaces. The principal example later in these notes is Wiener measure on

$$\{f \in C([0,1]) : f(0) = 0\},$$

the space of continuous paths starting at the origin.

4.2.4. *Baire-generic sets may have measure zero.* Recall that a set N in a topological space is *nowhere dense* if

$$\operatorname{int}(\overline{N}) = \emptyset.$$

A set is called *meagre* if it is a countable union of nowhere-dense sets, and it is called *residual* if its complement is meagre. A property is said to hold *Baire-generically* if it holds on a residual set.

If

$$G = \bigcap_{j=1}^{\infty} U_j,$$

where every U_j is open and dense, then

$$X \setminus G = \bigcup_{j=1}^{\infty} (X \setminus U_j)$$

is meagre. Moreover, if X is a complete metric space, the Baire category theorem says that G is dense. Thus a countable intersection of open dense sets is a particularly convenient example of a residual set.

Residuality is a topological notion of largeness. It need not agree with measure-theoretic largeness, as the following classical example shows.

Definition 4.10. *An irrational number $x \in \mathbb{R}$ is called a* Liouville number *if, for every $n \in \mathbb{N}$, there exist integers p and q, with $q \geq 2$, such that*

$$0 < \left| x - \frac{p}{q} \right| < \frac{1}{q^n}.$$

Thus a Liouville number can be approximated by rational numbers more accurately than any prescribed power of the denominator.

A standard example is

$$\lambda = \sum_{k=1}^{\infty} 10^{-k!} = 0.11000100000000000000000001\ldots.$$

Indeed, let

$$\lambda_N = \sum_{k=1}^{N} 10^{-k!} = \frac{p_N}{q_N}, \qquad q_N = 10^{N!}.$$

The tail satisfies

$$0 < \lambda - \lambda_N < 2 \cdot 10^{-(N+1)!}.$$

Given n, choose $N \geq \max\{n, 2\}$. Since

$$(N+1)! > nN!,$$

we obtain

$$0 < \left| \lambda - \frac{p_N}{q_N} \right| < 10^{-nN!} = \frac{1}{q_N^n}.$$

The decimal expansion of λ is not eventually periodic, so λ is irrational. Hence λ is a Liouville number. Liouville's theorem states, in fact, that every Liouville number is transcendental.

Let $\mathfrak{L}$ denote the set of all Liouville numbers.

Proposition 4.11. *The set $\mathfrak{L}$ is residual in $\mathbb{R}$, but it has Lebesgue measure zero.*

Proof. For each $n \in \mathbb{N}$, define

$$U_n = \bigcup_{q=2}^{\infty} \bigcup_{p \in \mathbb{Z}} \left(\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n} \right).$$

Each U_n is open. It is also dense, since it contains every rational number and $\mathbb{Q}$ is dense in $\mathbb{R}$. Moreover,

$$\mathfrak{L} = (\mathbb{R} \setminus \mathbb{Q}) \cap \bigcap_{n=1}^{\infty} U_n.$$

Indeed, for an irrational x, membership in every U_n is exactly the Liouville approximation property.

If $\mathbb{Q} = \{r_1, r_2, \ldots\}$, then

$$\mathbb{R} \setminus \mathbb{Q} = \bigcap_{j=1}^{\infty} (\mathbb{R} \setminus \{r_j\}),$$

and every $\mathbb{R} \setminus \{r_j\}$ is open and dense. Consequently, $\mathfrak{L}$ is a countable intersection of open dense sets. By the Baire category theorem, $\mathfrak{L}$ is dense and residual.

We now prove that $\mathfrak{L}$ has Lebesgue measure zero. Fix $M \geq 1$. If

$$\left(\frac{p}{q} - \frac{1}{q^n}, \frac{p}{q} + \frac{1}{q^n} \right)$$

meets $[-M, M]$, then

$$|p| \leq (M+1)q.$$

For each fixed q, there are therefore at most

$$2(M+1)q+1$$

relevant choices of p. Hence, for $n \geq 4$,

$$\begin{aligned} m^*(U_n \cap [-M,M]) &\leq \sum_{q=2}^{\infty} \big(2(M+1)q+1\big)\frac{2}{q^n} \\ &= 4(M+1)\sum_{q=2}^{\infty}\frac{1}{q^{n-1}} + 2\sum_{q=2}^{\infty}\frac{1}{q^n}. \end{aligned}$$

Both terms on the right tend to zero as $n \to \infty$. Since $\mathfrak{L} \subseteq U_n$ for every n, it follows that

$$m^*(\mathfrak{L} \cap [-M,M]) = 0.$$

Finally,

$$\mathfrak{L} = \bigcup_{M=1}^{\infty} \big(\mathfrak{L} \cap [-M,M]\big),$$

so countable subadditivity gives

$$m(\mathfrak{L}) = 0.$$

□

Thus $\mathfrak{L} \cap [0,1]$ is residual in $[0,1]$ but has Lebesgue measure zero. Its complement is meagre but has full Lebesgue measure. This shows that Baire category and measure give genuinely different notions of what it means for a set to be "large" or for a property to be "typical."

4.2.5. *Unbounded self-adjoint operators and spectral measures.* Many important operators on L^2 are unbounded. For such an operator, its domain is part of its definition. Let

$$\mathcal{H} := L^2(X,\mu)$$

and let

$$A : \mathcal{D}(A) \subseteq \mathcal{H} \longrightarrow \mathcal{H}$$

be a densely defined self-adjoint operator. A complex number z belongs to the resolvent set $\rho(A)$ if

$$A - zI : \mathcal{D}(A) \longrightarrow \mathcal{H}$$

is bijective and its inverse is bounded as an operator on $\mathcal{H}$. The spectrum of A is

$$\sigma(A) := \mathbb{C} \setminus \rho(A).$$

Self-adjointness implies that

$$\sigma(A) \subseteq \mathbb{R}.$$

The spectral theorem provides a projection-valued Borel measure E_A on $\mathbb{R}$ such that, symbolically,

$$A = \int_{\mathbb{R}} t \, dE_A(t).$$

For $f \in \mathcal{H}$, define the scalar spectral measure

$$\mu_f(B) := \langle E_A(B) f, f \rangle, \qquad B \in \mathcal{B}(\mathbb{R}).$$

This is a finite positive measure satisfying

$$\mu_f(\mathbb{R}) = \|f\|_2^2.$$

In particular, μ_f is a probability measure when $\|f\|_2 = 1$.

The domain and the action of A are encoded by these measures:

$$\mathcal{D}(A) = \left\{ f \in \mathcal{H} : \int_{\mathbb{R}} t^2 \, d\mu_f(t) < \infty \right\},$$

and, for $f \in \mathcal{D}(A)$,

$$\|Af\|_2^2 = \int_{\mathbb{R}} t^2 \, d\mu_f(t).$$

Moreover, for every bounded Borel function φ,

$$\langle \varphi(A) f, f \rangle = \int_{\mathbb{R}} \varphi(t) \, d\mu_f(t).$$

Thus the spectral theorem converts questions about a self-adjoint operator into questions about measures on the real line.

Multiplication operators and essential range. The fundamental model is a multiplication operator. Let (X, μ) be a σ-finite measure space, and let $h : X \to \mathbb{R}$ be measurable and finite almost everywhere. Define

$$(M_h f)(x) := h(x) f(x)$$

on the maximal domain

$$\mathcal{D}(M_h) = \{ f \in L^2(X, \mu) : hf \in L^2(X, \mu) \}.$$

This domain is dense. Indeed, for $f \in L^2(X, \mu)$, set

$$f_n := f \chi_{\{|h| \le n\}}.$$

Then $f_n \in \mathcal{D}(M_h)$, since $|hf_n| \le n|f|$, and, because h is finite almost everywhere,

$$\|f - f_n\|_2^2 = \int_{\{|h| > n\}} |f|^2 \, d\mu \longrightarrow 0.$$

The operator M_h is self-adjoint. It is symmetric because h is real-valued:

$$\langle M_h f, g \rangle = \int_X hf\overline{g} \, d\mu = \int_X f\overline{hg} \, d\mu = \langle f, M_h g \rangle$$

for $f, g \in \mathcal{D}(M_h)$.

Conversely, suppose that $g \in \mathcal{D}(M_h^*)$ and write

$$M_h^* g = k.$$

Then

$$\langle M_h f, g \rangle = \langle f, k \rangle \qquad \text{for every } f \in \mathcal{D}(M_h).$$

Let

$$E_n := \{|h| \le n\}.$$

Every $f \in L^2(X, \mu)$ supported in E_n belongs to $\mathcal{D}(M_h)$. Consequently,

$$\int_{E_n} f \, \overline{(hg - k)} \, d\mu = 0$$

for every such f. Since

$$(hg - k)\chi_{E_n} \in L^2(X, \mu),$$

we may take $f = (hg - k)\chi_{E_n}$ and obtain

$$\int_{E_n} |hg - k|^2 \, d\mu = 0.$$

Thus $hg = k$ almost everywhere on E_n. Since $E_n \nearrow X$ up to a null set,

$$hg = k \qquad \text{almost everywhere on } X.$$

It follows that $hg \in L^2(X,\mu)$, so $g \in \mathcal{D}(M_h)$ and $M_h^* g = M_h g$. Hence

$$M_h^* = M_h.$$

The following is the main reason of interest in multiplication operators:

Theorem 4.12 (Unbounded self-adjoint)**.** *If A is an unbounded, self-adjoint operator on a separable Hilbert space H, there is a measure space (Ω,μ), a unitary map $W : L^2(\Omega, d\mu) \to H$, and a real-valued measurable function a on Ω such that*

$$W^{-1}AW\ f(x) = a(x)f(x), \quad Wf \in \mathcal{D}(A). \tag{43}$$

In this situation, given $f \in L^2(\Omega, d\mu)$, Wf belongs to $\mathcal{D}(A)$ iff $af \in L^2(\Omega, d\mu)$.

Suppose X is a measure space, and $M_f : L^2(X) \to L^2(X)$ is the multiplication operator by a function $f \in L^\infty_{\text{loc}}$. Without loss of generality, we can assume that μ is a finite measure.

Lemma 4.13 (Rudin's finite-measure reduction)**.** *Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space. Then there exists a measurable function $w : X \to (0,1)$ such that*

$$w \in L^1(\mu).$$

Consequently, the measure

$$\nu(A) := \int_A w\, d\mu, \qquad A \in \mathcal{M},$$

is finite and has exactly the same null sets as μ.

Proof. By σ-finiteness, there are measurable sets $E_n \subseteq X$ such that

$$X = \bigcup_{n=1}^{\infty} E_n, \qquad \mu(E_n) < \infty.$$

Define

$$w(x) := \sum_{n=1}^{\infty} \frac{2^{-n-1}}{1+\mu(E_n)} \chi_{E_n}(x).$$

Since the sets E_n cover X, at least one term in this sum is positive at every $x \in X$. Hence $w(x) > 0$ for every x. Moreover,

$$w(x) \le \sum_{n=1}^{\infty} 2^{-n-1} < 1.$$

By the Monotone Convergence Theorem,

$$\int_X w\, d\mu = \sum_{n=1}^{\infty} \frac{2^{-n-1}\mu(E_n)}{1+\mu(E_n)} \le \sum_{n=1}^{\infty} 2^{-n-1} < \infty.$$

Thus $w \in L^1(\mu)$ and

$$\nu(X) = \int_X w\, d\mu < \infty.$$

If $\mu(A) = 0$, then clearly $\nu(A) = 0$. Conversely, suppose that $\nu(A) = 0$. Since

$$A = \bigcup_{k=1}^{\infty} \left(A \cap \{w \ge 1/k\}\right),$$

if $\mu(A) > 0$, then for some k one would have

$$\mu\left(A \cap \{w \ge 1/k\}\right) > 0.$$

But then

$$\nu(A) \geq \int_{A\cap\{w\geq 1/k\}} w\, d\mu \geq \frac{1}{k}\,\mu\left(A\cap\{w\geq 1/k\}\right) > 0,$$

a contradiction. Therefore $\mu(A) = 0$, and μ and ν have the same null sets. □

The preceding lemma shows that a multiplication model over a σ-finite measure space can always be replaced by an equivalent model over a finite measure space. More precisely, let $f : X \to \mathbb{R}$ be measurable and let

$$M_f^\mu g = fg, \qquad \mathcal{D}(M_f^\mu) = \{g \in L^2(\mu) : fg \in L^2(\mu)\}.$$

Define

$$U : L^2(\mu) \longrightarrow L^2(\nu), \qquad Ug = w^{-1/2}g.$$

Then

$$\|Ug\|^2_{L^2(\nu)} = \int_X |w^{-1/2}g|^2 w\, d\mu = \int_X |g|^2\, d\mu,$$

so U is an isometry. It is surjective, with inverse

$$U^{-1}h = w^{1/2}h,$$

and hence is unitary. Furthermore,

$$g \in \mathcal{D}(M_f^\mu) \iff Ug \in \mathcal{D}(M_f^\nu),$$

and, for every $g \in \mathcal{D}(M_f^\mu)$,

$$UM_f^\mu g = M_f^\nu Ug.$$

Thus M_f^μ and M_f^ν are unitarily equivalent.

Since μ and ν have the same null sets, the essential range of f and the condition

$$\mu\big(f^{-1}(\{\lambda\})\big) > 0$$

are unchanged when μ is replaced by ν. We may therefore assume, when studying a multiplication representation of a self-adjoint operator, that the underlying measure is finite. After a harmless normalization, it may be taken to be a probability measure.

We have the following result, which complements the spectral theorem. First, we define the essential range of a function.

Definition 4.14. *ess ran f is the set of all those complex numbers z such that the inverse image of any ε-neighbourhood of z under f has positive measure.*

Theorem 4.15.

$$\mathrm{Spec}(M_f) = \textit{ess ran } f$$
$$\mathrm{Spec}_p(M_f) = \{\lambda \in \mathbb{C} : \mu\left(f^{-1}\{\lambda\}\right) > 0\}.$$

Proof. Pick a point λ outside ess ran f. We would like to show that λ is in the resolvent. By definition, $\frac{1}{f-\lambda} \in L^\infty_{\mathrm{loc}}$. It boils down to proving that if $(M_f-\lambda)^{-1}u = v$, then $\|u\| = \|(f-\lambda)v\| \geq c\|v\|$, which follows from the definition of essential range. This shows that $(M_f-\lambda)^{-1}$ is bounded. Clearly, it is the inverse of $M_f - \lambda$.

Conversely, if λ is in the essential range, then there are two cases.

Case I: $f^{-1}(\lambda)$ has positive measure. Then we want to check that λ is in the point spectrum of M_f. Observe that saying $\lambda \in \mathrm{Spec}_p M_f$ is equivalent to the existence of $\varphi \in L^2(X)$ such that $(f(x)-\lambda)\varphi(x) = 0$ a.e. Since $S := f^{-1}(\lambda)$ has positive measure, consider $\varphi = \chi_S$, and we are done.

Case II: $f^{-1}(\lambda)$ has zero measure. Clearly, $f - \lambda \in L^\infty_{\mathrm{loc}}$. For any $m \in \mathbb{N}$ denote $S_m := \{x \in X : |f(x) - \lambda| < 2^{-m}\}$, which have strictly positive but finite measure. Clearly, one has

$$\|(M_f - \lambda)\chi_{S_m}\|^2 = \int_{S_m} |f(x) - \lambda|^2 |\chi_{S_m}|^2 \leq 2^{-2m} \|\chi_{S_m}\|^2,$$

which proves that $(M_f - \lambda)^{-1}$ cannot be bounded. □

The most important example of the above theorem is that of the Laplacian on $L^2(\mathbb{R}^n)$. Since the Laplacian is unitarily (via the Fourier transform) equivalent to multiplication by $\|x\|^2$, it is clear that $\mathrm{Spec}(-\Delta) = [0, \infty)$, and there are no eigenvalues.

The bounded Borel functional calculus for M_h is explicit:

$$\varphi(M_h) f = (\varphi \circ h) f.$$

Taking $\varphi = \chi_B$, we obtain

$$(E_{M_h}(B) f)(x) = \chi_{h^{-1}(B)}(x) f(x).$$

Consequently,

$$\mu_f(B) = \int_{h^{-1}(B)} |f|^2 \, d\mu = h_\#\big(|f|^2 \mu\big)(B).$$

Thus the spectral measure of f is the pushforward, under h, of the measure $|f|^2 \, d\mu$.

The three spectral types. Write m for Lebesgue measure on $\mathbb{R}$. Every scalar spectral measure has a unique decomposition

$$\mu_f = \mu_f^{\mathrm{pp}} + \mu_f^{\mathrm{ac}} + \mu_f^{\mathrm{sc}},$$

where

(1) μ_f^{pp} is purely atomic;
(2) $\mu_f^{\mathrm{ac}} \ll m$, and hence
$$d\mu_f^{\mathrm{ac}}(t) = w_f(t) \, dt$$
for some $w_f \in L^1(\mathbb{R})$;
(3) $\mu_f^{\mathrm{sc}} \perp m$, but
$$\mu_f^{\mathrm{sc}}(\{\lambda\}) = 0 \qquad \text{for every } \lambda \in \mathbb{R}.$$

The abbreviation "pp" stands for *pure point*, while "sc" stands for *singular continuous*. A singular continuous measure is concentrated on a Lebesgue-null set but has no atoms; Cantor measure is the basic example.

The decomposition of the scalar spectral measures gives canonical mutually orthogonal reducing subspaces

$$\mathcal{H} = \mathcal{H}_{\mathrm{pp}} \oplus \mathcal{H}_{\mathrm{ac}} \oplus \mathcal{H}_{\mathrm{sc}},$$

where

$$\begin{aligned}
\mathcal{H}_{\mathrm{pp}} &= \{f \in \mathcal{H} : \mu_f \text{ is purely atomic}\} \\
&= \overline{\mathrm{span}}\{\ker(A - \lambda I) : \lambda \in \mathbb{R}\}, \\
\mathcal{H}_{\mathrm{ac}} &= \{f \in \mathcal{H} : \mu_f \ll m\}, \\
\mathcal{H}_{\mathrm{sc}} &= \big\{f \in \mathcal{H} : \mu_f \perp m \text{ and } \mu_f(\{\lambda\}) = 0 \text{ for every } \lambda \in \mathbb{R}\big\}.
\end{aligned}$$

Because these subspaces reduce A,

$$\mathcal{D}(A) = \bigoplus_{\tau \in \{\mathrm{pp,ac,sc}\}} \big(\mathcal{D}(A) \cap \mathcal{H}_\tau\big).$$

For $\tau \in \{\mathrm{pp}, \mathrm{ac}, \mathrm{sc}\}$, define

$$A_\tau := A \restriction_{\mathcal{D}(A) \cap \mathcal{H}_\tau}.$$

Each A_τ is self-adjoint on $\mathcal{H}_\tau$, and

$$A = A_{\mathrm{pp}} \oplus A_{\mathrm{ac}} \oplus A_{\mathrm{sc}}.$$

We write

$$\sigma_\tau(A) := \sigma(A_\tau).$$

These are, respectively, the pure-point, absolutely continuous and singular continuous parts of the spectrum, and

$$\sigma(A) = \sigma_{\mathrm{pp}}(A) \cup \sigma_{\mathrm{ac}}(A) \cup \sigma_{\mathrm{sc}}(A).$$

This is an operator decomposition, not in general a disjoint set-theoretic partition: the three closed sets on the right may overlap.

The point spectrum is the set of actual eigenvalues,

$$\sigma_{\mathrm{p}}(A) = \{\lambda \in \mathbb{R} : \ker(A - \lambda I) \neq \{0\}\}.$$

It should not be confused with $\sigma_{\mathrm{pp}}(A)$. In fact,

$$\sigma_{\mathrm{pp}}(A) = \overline{\sigma_{\mathrm{p}}(A)}.$$

Thus a limit point of eigenvalues can belong to $\sigma_{\mathrm{pp}}(A)$ without itself being an eigenvalue.

Multiplication operators give concrete examples of all three types.

(1) On

$$\ell^2(\{1, 2, \dots\}) = L^2(\{1, 2, \dots\}, \text{counting measure}),$$

define

$$(Aa)_n = n a_n$$

on

$$\mathcal{D}(A) = \left\{ a : \sum_{n \geq 1} n^2 |a_n|^2 < \infty \right\}.$$

Then

$$\sigma(A) = \{1, 2, \dots\},$$

and A has pure-point spectral type.

(2) Multiplication by x on $L^2(\mathbb{R}, dx)$ has

$$\sigma(M_x) = \mathbb{R}.$$

Moreover,

$$\mu_f(B) = \int_B |f(x)|^2 \, dx,$$

so every spectral measure is absolutely continuous. Thus M_x has purely absolutely continuous spectrum.

Similarly, let

$$H_0 = -\frac{d^2}{dx^2}, \qquad \mathcal{D}(H_0) = H^2(\mathbb{R}).$$

Under the usual unitary Fourier transform, H_0 becomes multiplication by ξ^2. Consequently,

$$\sigma(H_0) = [0, \infty),$$

and H_0 has purely absolutely continuous spectrum.

(3) Let ν_C be Cantor measure, and consider multiplication by x on $L^2([0,1],\nu_C)$. Since

$$\operatorname{supp}\nu_C = C,$$

where C is the Cantor set,

$$\sigma(M_x) = \operatorname{ess\,ran}_{\nu_C}(x) = C.$$

For every $f \in L^2([0,1],\nu_C)$,

$$\mu_f(B) = \int_B |f(x)|^2\, d\nu_C(x).$$

Thus

$$\mu_f \ll \nu_C \perp m.$$

Moreover, since ν_C has no atoms,

$$\mu_f(\{\lambda\}) = 0 \qquad \text{for every } \lambda \in \mathbb{R}.$$

Hence every non-zero scalar spectral measure is singular continuous, and M_x has purely singular continuous spectrum.

The spectrum as a closed set does not determine the spectral type. For example, multiplication by x on $L^2([0,1],dx)$ has spectrum $[0,1]$ and purely absolutely continuous spectral type. On the other hand, enumerate $\mathbb{Q}\cap[0,1]$ as $\{q_1,q_2,\dots\}$ and define

$$De_n = q_n e_n$$

on $\ell^2(\mathbb{N})$. Then

$$\sigma(D) = \overline{\{q_1,q_2,\dots\}} = [0,1],$$

but D has pure-point spectral type. Thus two operators can have exactly the same spectrum as a set while having completely different spectral measures.

Remark 4.16 (Simon's Wonderland theorem)**.** *Let $\mathfrak{X}$ be a regular metric space of self-adjoint operators on a fixed separable Hilbert space. This means that $\mathfrak{X}$ is complete and that convergence in its metric implies strong-resolvent convergence. Let $J \subseteq \mathbb{R}$ be an open interval. Suppose that each of the sets*

$$\{T \in \mathfrak{X} : \sigma_{\mathrm{pp}}(T) \cap J = \emptyset\},$$

$$\{T \in \mathfrak{X} : \sigma_{\mathrm{ac}}(T) \cap J = \emptyset\},$$

and

$$\{T \in \mathfrak{X} : J \subseteq \sigma(T)\}$$

is dense in $\mathfrak{X}$. Then their intersection is a dense G_δ subset of $\mathfrak{X}$. For every operator in this intersection,

$$J \subseteq \sigma_{\mathrm{sc}}(T).$$

Thus, under these density hypotheses, purely singular continuous spectrum on J is generic in the sense of Baire category.[6]

This is a category statement, not an almost-everywhere statement with respect to a probability measure on $\mathfrak{X}$. The three density hypotheses are essential.

4.3. **Classical geometric measures: Besicovitch, De Giorgi, and Federer.** Lebesgue measure is not the only reasonable way of measuring a subset of Euclidean space. Different measures can detect different aspects of the same set: its length, its visibility under projection, its curvature, or the size of its measure-theoretic boundary. The examples below were among the starting points of geometric measure theory.

[6] B. Simon, *Operators with singular continuous spectrum. I. General operators*, Ann. of Math. (2) **141** (1995), no. 1, 131–145, Theorem 2.1.

4.3.1. *Hausdorff measure and integral-geometric measure.* We use $\mathcal{H}^1$ for one-dimensional Hausdorff measure, which will be defined formally later. A Borel set $E \subseteq \mathbb{R}^2$ is called *countably* 1*-rectifiable* if, up to a set of $\mathcal{H}^1$-measure zero, it is contained in a countable union of Lipschitz curves. At the opposite extreme, E is called *purely* 1*-unrectifiable* if

$$\mathcal{H}^1(E \cap \Gamma) = 0$$

for every Lipschitz curve Γ.

There is a natural measure which detects this distinction. Parametrize an unoriented line by

$$L_{\theta,p} = \{x \in \mathbb{R}^2 : x \cdot (\cos\theta, \sin\theta) = p\}, \qquad 0 \le \theta < \pi, \quad p \in \mathbb{R}.$$

For a Borel set $A \subseteq \mathbb{R}^2$, define its *integral-geometric measure* by

$$\mathcal{I}^1(A) = \frac{1}{2}\int_0^\pi \int_{\mathbb{R}} \#(A \cap L_{\theta,p})\, dp\, d\theta.$$

This is an extended Borel measure. Indeed, if $\{A_j\}$ is a pairwise disjoint collection, then for every line $L_{\theta,p}$,

$$\#\left(\left(\bigcup_j A_j\right) \cap L_{\theta,p}\right) = \sum_j \#(A_j \cap L_{\theta,p}).$$

Integrating and applying the Monotone Convergence Theorem gives countable additivity.

The measure $\mathcal{I}^1$ is not locally finite: every open ball has infinite $\mathcal{I}^1$-measure, since its intersection with almost every line passing through it contains infinitely many points. Nevertheless, $\mathcal{I}^1$ is finite and useful on many one-dimensional sets.

For a countably 1-rectifiable set E, the Crofton formula gives

$$\mathcal{I}^1(E) = \mathcal{H}^1(E).$$

For purely 1-unrectifiable sets, however, the situation is completely different. The Besicovitch projection theorem states that if

$$E \subseteq \mathbb{R}^2, \qquad \mathcal{H}^1(E) < \infty,$$

is purely 1-unrectifiable, then

$$m_1(\pi_\theta(E)) = 0$$

for almost every orthogonal projection $\pi_\theta : \mathbb{R}^2 \to \mathbb{R}$. Consequently,

$$\mathcal{I}^1(E) = 0.$$

Thus there exist Borel sets for which

$$0 < \mathcal{H}^1(E) < \infty, \qquad \mathcal{I}^1(E) = 0.$$

One measure assigns positive length to the set, while the other sees nothing because the set is invisible to almost every family of parallel lines.[7]

The same phenomenon can be expressed using pushforwards. Set

$$\mu = \mathcal{H}^1 \lfloor_E$$

and, for each direction θ, define

$$\mu_\theta = (\pi_\theta)_\# \mu.$$

If E is purely 1-unrectifiable, then for almost every θ the measure μ_θ has positive total mass but is supported on the Lebesgue-null set $\pi_\theta(E)$. Hence

$$\mu_\theta \perp m_1 \qquad \text{for almost every } \theta.$$

[7]See A. S. Besicovitch, *On the fundamental geometrical properties of linearly measurable plane sets of points (III)*, Math. Ann. **116** (1939), 349–357, and H. Federer, *Some integralgeometric theorems*, Trans. Amer. Math. Soc. **77** (1954), 238–261.

A finite length measure can therefore cast a singular shadow in almost every direction.

A concrete model is obtained from the one-quarter Cantor set:

$$K = \left\{ \sum_{j=1}^{\infty} \frac{3\varepsilon_j}{4^j} : \varepsilon_j \in \{0,1\} \right\} \subseteq [0,1]$$

Let

$$E := K \times K \subseteq \mathbb{R}^2.$$

At the n-th stage of its construction, E is contained in 4^n squares of side length 4^{-n}, one at each corner of every square retained at the preceding stage. It is therefore called the *four-corner Cantor set*.

Proposition 4.17. *The four-corner Cantor set satisfies*

$$0 < \mathcal{H}^1(E) < \infty.$$

Moreover, it is purely 1*-unrectifiable: for every rectifiable curve* $\Gamma \subseteq \mathbb{R}^2$*,*

$$\mathcal{H}^1(E \cap \Gamma) = 0.$$

Finally,

$$\mathcal{I}^1(E) = 0.$$

Proof. The 4^n squares occurring at the n-th stage have diameter $\sqrt{2}\,4^{-n}$. Consequently,

$$\mathcal{H}^1(E) \le \liminf_{n\to\infty} 4^n \sqrt{2}\, 4^{-n} = \sqrt{2},$$

so $\mathcal{H}^1(E) < \infty$.

To prove positivity, let μ be the natural probability measure on E which assigns mass 4^{-n} to every square occurring at the n-th stage[8]. If

$$4^{-n} \le r < 4^{-n+1},$$

then a ball of radius r can meet at most a fixed number of the n-th generation squares. Indeed, distinct squares of that generation are separated in at least one coordinate by a distance comparable to 4^{-n}. Hence

$$\mu(B(x,r)) \le C 4^{-n} \le Cr$$

for some constant C independent of x and r.

Let μ^* denote the outer measure induced by μ. If $U \subseteq \mathbb{R}^2$ meets E, choose $x \in U \cap E$. Then

$$U \cap E \subseteq B(x, \operatorname{diam}(U)),$$

and therefore

$$\mu^*(U) \le C \operatorname{diam}(U).$$

Thus, for every covering $E \subseteq \bigcup_j U_j$,

$$1 = \mu(E) \le \sum_j \mu^*(U_j) \le C \sum_j \operatorname{diam}(U_j).$$

Taking the infimum over all such coverings gives

$$\mathcal{H}^1(E) \ge C^{-1} > 0.$$

We next prove pure unrectifiability. Notice first that K has Lebesgue measure zero, since its n-th approximation consists of 2^n intervals of length 4^{-n} and hence has total length

$$2^n 4^{-n} = 2^{-n}.$$

[8]Again, the usual 0 or 1 chooses the left or right square in the x-coordinate, and same for the y-coordinate. Each square at the n-th stage has then taken n prescribed pairs of choices.

Let

$$\gamma(t) = (x(t), y(t)), \qquad t \in [a, b],$$

be a Lipschitz parametrisation of a rectifiable curve, and set

$$A := \{t \in [a, b] : \gamma(t) \in E\}.$$

Since $x(A) \subseteq K$, the one-dimensional area formula gives

$$\int_A |x'(t)|\, dt = \int_{\mathbb{R}} \#\{t \in A : x(t) = s\}\, ds = 0.$$

Similarly, because $y(A) \subseteq K$,

$$\int_A |y'(t)|\, dt = 0.$$

Applying the area formula to γ itself, we obtain

$$\begin{aligned} \mathcal{H}^1\big(E \cap \gamma([a, b])\big) &= \mathcal{H}^1(\gamma(A)) \\ &\leq \int_A |\gamma'(t)|\, dt \\ &\leq \int_A \big(|x'(t)| + |y'(t)|\big)\, dt \\ &= 0. \end{aligned}$$

Thus E is purely 1-unrectifiable.

The final assertion is an application of the Besicovitch–Federer projection theorem. It states that if $A \subseteq \mathbb{R}^2$ is purely 1-unrectifiable and $\mathcal{H}^1(A) < \infty$, then

$$\mathcal{L}^1\big(\pi_\theta(A)\big) = 0$$

for almost every $\theta \in [0, \pi)$, where

$$\pi_\theta(z) = z \cdot (\cos\theta, \sin\theta).$$

For such a θ, the function

$$p \longmapsto \#(E \cap L_{\theta,p})$$

vanishes outside the null set $\pi_\theta(E)$. Therefore

$$\int_{\mathbb{R}} \#(E \cap L_{\theta,p})\, dp = 0$$

for almost every θ. Consequently,

$$\mathcal{I}^1(E) = \frac{1}{2} \int_0^\pi \int_{\mathbb{R}} \#(E \cap L_{\theta,p})\, dp\, d\theta = 0.$$

□

4.3.2. *Kakeya sets and mutually singular length measures.* A direction in the plane will be represented by an angle $\theta \in [0, \pi)$, since the angles θ and $\theta + \pi$ determine the same unoriented direction. Write

$$e_\theta = (\cos\theta, \sin\theta).$$

Definition 4.18. *A set $\mathscr{K} \subseteq \mathbb{R}^2$ is called a* Kakeya set, *or a* Besicovitch set, *if, for every $\theta \in [0, \pi)$, there exists $x_\theta \in \mathbb{R}^2$ such that*

$$x_\theta + [0, 1]e_\theta \subseteq \mathscr{K}.$$

Thus $\mathscr{K}$ contains a unit line segment in every direction.

Terminology varies slightly here. The original Kakeya needle problem asks for a set inside which a unit segment can be moved continuously through all directions. In this subsection, we use "Kakeya set" in the now standard harmonic-analytic sense of a set which merely contains a unit segment in every direction.

The following theorem is one of the most striking early discoveries of geometric measure theory.

Theorem 4.19 (Besicovitch)**.** *There exists a compact Kakeya set $\mathscr{K} \subseteq \mathbb{R}^2$ such that*

$$m_2(\mathscr{K}) = 0.$$

Besicovitch's original construction was subsequently simplified by Perron.[9]

For each $(u, v) \in E = C \times C$, consider the line segment

$$\ell_{u,v} = \big\{ \big(x, u + x(u + 2v)\big) : 0 \le x \le 1 \big\}.$$

Its slope is

$$s(u, v) = u + 2v.$$

Define

$$B_0 = \bigcup_{(u,v)\in E} \ell_{u,v} = \big\{ \big(x, u + x(u + 2v)\big) : 0 \le x \le 1,\ (u, v) \in E \big\}.$$

This set is compact because it is the continuous image of the compact set $[0, 1] \times E$.

Since

$$\{u + 2v : (u, v) \in E\} = C + 2C = [0, 3],$$

the set B_0 contains a segment of every slope $s \in [0, 3]$. The segment parametrised by

$$x \longmapsto (x, u + sx), \qquad 0 \le x \le 1,$$

has length $\sqrt{1 + s^2}$. An exactly unit subsegment is obtained by restricting the parameter to

$$0 \le x \le \frac{1}{\sqrt{1 + s^2}}.$$

Thus B_0 contains a unit segment in every direction whose angle lies between

$$0 \qquad \text{and} \qquad \beta := \arctan 3.$$

We now show that B_0 has planar Lebesgue measure zero. Fix $x \in [0, 1]$. Its vertical section is

$$\begin{aligned}(B_0)_x &= \{u + x(u + 2v) : (u, v) \in E\} \\ &= (1 + x)C + 2xC.\end{aligned}$$

Set

$$w_x = (1 + x, 2x)$$

and let

$$\alpha(x) = \arg(w_x) = \arctan\left(\frac{2x}{1 + x}\right).$$

If $\pi_{\alpha(x)}$ denotes orthogonal projection onto the line in direction w_x, identified with a scalar projection onto $\mathbb{R}$, then

$$(B_0)_x = |w_x|\, \pi_{\alpha(x)}(E).$$

Let $\mathcal{N} \subset [0, \pi)$ denote the exceptional set of directions in the Besicovitch–Federer projection theorem. Thus

$$\mathcal{L}^1\big(\pi_\theta(E)\big) = 0 \qquad \text{whenever } \theta \notin \mathcal{N},$$

and $\mathcal{L}^1(\mathcal{N}) = 0$. We must check that the direction $\alpha(x)$ belongs to $\mathcal{N}$ for only a null set of values of x.

[9] See A. S. Besicovitch, *On Kakeya's problem and a similar one*, Math. Z. **27** (1928), 312–320, and O. Perron, *Über einen Satz von Besicovitsch*, Math. Z. **28** (1928), 383–386.

Recall that

$$\alpha(x) = \arctan\left(\frac{2x}{1+x}\right), \qquad 0 \le x \le 1.$$

Differentiating gives

$$\alpha'(x) = \frac{2}{(1+x)^2 + 4x^2} = \frac{2}{1 + 2x + 5x^2}.$$

The denominator is increasing on $[0,1]$ and takes values between 1 and 8. Consequently,

$$\frac{1}{4} \le \alpha'(x) \le 2 \qquad \text{for every } x \in [0,1].$$

In particular, α is strictly increasing. Moreover, the Mean Value Theorem gives, for every $x, y \in [0,1]$,

$$\frac{1}{4}|x-y| \le |\alpha(x) - \alpha(y)| \le 2|x-y|.$$

Thus α is bi-Lipschitz from $[0,1]$ onto its image

$$J := \alpha([0,1]),$$

and its inverse $\alpha^{-1} : J \to [0,1]$ is Lipschitz with constant at most 4.

We now show explicitly that

$$\alpha^{-1}(\mathcal{N}) = \{x \in [0,1] : \alpha(x) \in \mathcal{N}\}$$

has measure zero. Fix $\varepsilon > 0$. Since $\mathcal{L}^1(\mathcal{N} \cap J) = 0$, there are intervals $\{I_k\}_{k\ge1}$ such that

$$\mathcal{N} \cap J \subseteq \bigcup_{k\ge1} I_k, \qquad \sum_{k\ge1} |I_k| < \frac{\varepsilon}{4}.$$

Because α^{-1} is 4-Lipschitz, each set $\alpha^{-1}(I_k \cap J)$ has diameter at most $4|I_k|$. Therefore

$$\mathcal{L}^1\big(\alpha^{-1}(\mathcal{N})\big) \le \sum_{k\ge1} \operatorname{diam}\big(\alpha^{-1}(I_k \cap J)\big) \le 4 \sum_{k\ge1} |I_k| < \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary,

$$\mathcal{L}^1\big(\alpha^{-1}(\mathcal{N})\big) = 0.$$

Hence, for almost every $x \in [0,1]$, the direction $\alpha(x)$ is not exceptional, and therefore

$$\mathcal{L}^1\big(\pi_{\alpha(x)}(E)\big) = 0.$$

As observed above, $(B_0)_x$ is a translate and a scalar multiple of $\pi_{\alpha(x)}(E)$. Translations and scalar dilations preserve Lebesgue-null sets, so

$$\mathcal{L}^1((B_0)_x) = 0$$

for almost every $x \in [0,1]$.

Since B_0 is compact and hence Borel measurable, Fubini's theorem gives

$$m_2(B_0) = \int_{\mathbb{R}} \mathcal{L}^1((B_0)_x)\, dx = \int_0^1 \mathcal{L}^1((B_0)_x)\, dx = 0.$$

It remains only to cover the directions not represented by slopes in $[0,3]$. Let R_γ denote rotation through the angle γ, and set

$$\mathscr{K} = B_0 \cup R_\beta(B_0) \cup R_{2\beta}(B_0), \qquad \beta = \arctan 3.$$

The three copies contain unit segments in the angular intervals

$$[0,\beta], \qquad [\beta, 2\beta], \qquad [2\beta, 3\beta],$$

respectively. Since

$$\beta = \arctan 3 > \arctan(\sqrt{3}) = \frac{\pi}{3},$$

we have

$$3\beta > \pi.$$

Hence these intervals cover all unoriented directions. Moreover, $\mathscr{K}$ is compact and

$$m_2(\mathscr{K}) = 0,$$

because it is a finite union of rotated copies of the compact null set B_0. This proves Besicovitch's theorem.

4.3.3. *Federer's curvature measures.* Let $K \subseteq \mathbb{R}^n$ be compact. Its *reach* is defined by

$$\operatorname{reach}(K) = \sup\left\{r > 0 : d(x,K) < r \Rightarrow x \text{ has a unique nearest point in } K\right\},$$

where the implication is required for every $x \in \mathbb{R}^n$ and $\sup\emptyset = 0$. Convex sets have infinite reach, while compact $C^{1,1}$ submanifolds without boundary have positive reach. If K has positive reach and

$$0 < r < \operatorname{reach}(K),$$

let

$$K_r = \{x \in \mathbb{R}^n : d(x,K) \le r\}$$

and let $p_K : K_r \to K$ denote the nearest-point projection. Federer proved that there are unique finite signed Borel measures

$$C_0(K,\cdot), \dots, C_n(K,\cdot)$$

such that, for every Borel set $A \subseteq K$,

$$m_n\big(K_r \cap p_K^{-1}(A)\big) = \sum_{j=0}^{n} \omega_{n-j} r^{n-j} C_j(K,A),$$

where ω_k denotes the volume of the unit ball in $\mathbb{R}^k$ and $\omega_0 = 1$. These are the *curvature measures* of K. The formula says that the coefficients in the polynomial expansion of the volume of a tubular neighbourhood are themselves measures. They record where the different contributions to the tube volume originate. For convex sets the curvature measures are positive; for general sets of positive reach they may be signed. Consider a convex polygon $P \subseteq \mathbb{R}^2$. Its curvature measures are

$$C_2(P,\cdot) = m_2 \mathbin{\lfloor} P, \qquad C_1(P,\cdot) = \frac{1}{2}\mathcal{H}^1 \mathbin{\lfloor} \partial P,$$

and

$$C_0(P,\cdot) = \sum_{v} \frac{\alpha_v}{2\pi}\delta_v,$$

where the sum runs over the vertices of P and α_v is the exterior angle at v. Thus curvature, which is represented by a function on a smooth curve, becomes an atomic measure concentrated at the corners of a polygon. Since

$$\sum_{v} \alpha_v = 2\pi,$$

we obtain

$$C_0(P,P) = 1.$$

Signed curvature already appears for an elementary nonconvex set with smooth boundary. Let

$$A_{a,b} = \{x \in \mathbb{R}^2 : a \le |x| \le b\}, \qquad 0 < a < b,$$

so that $\operatorname{reach}(A_{a,b}) = a$. For $0 < r < a$,

$$\begin{aligned} m_2\big((A_{a,b})_r\big) &= \pi(b+r)^2 - \pi(a-r)^2 \\ &= \pi(b^2 - a^2) + 2\pi(a+b)r. \end{aligned}$$

There is no quadratic term, so

$$C_0(A_{a,b}, A_{a,b}) = 0.$$

The two boundary circles make opposite contributions, as the localized form of the expansion shows. Writing $S_\rho = \{|x| = \rho\}$, points of the tube nearest to the outer circle are those with $b \le |x| \le b + r$, and points nearest to the inner circle are those with $a - r \le |x| \le a$, so

$$\begin{aligned} m_2\big((A_{a,b})_r \cap p^{-1}(S_b)\big) &= 2\pi br + \pi r^2, \\ m_2\big((A_{a,b})_r \cap p^{-1}(S_a)\big) &= 2\pi ar - \pi r^2, \end{aligned}$$

whence $C_0(A_{a,b}, S_b) = 1$ and $C_0(A_{a,b}, S_a) = -1$. More generally, for compact sets of positive reach, the total mass of C_0 equals the Euler characteristic. The annulus has Euler characteristic zero, whereas a convex body has Euler characteristic one. Federer's curvature measures therefore provide a natural transition from positive measures to signed measures.[10]

4.3.4. *De Giorgi's perimeter measure.* The topological boundary of a measurable set is often much too large to represent its surface area. De Giorgi's definition avoids this problem.

Let $E \subseteq \mathbb{R}^n$ be measurable and let $U \subseteq \mathbb{R}^n$ be open. Define

$$P(E, U) = \sup\left\{ \int_E \operatorname{div} \varphi \, dx : \varphi \in C_c^1(U; \mathbb{R}^n),\ |\varphi| \le 1 \right\}.$$

If $P(E, U) < \infty$ for every relatively compact open set U, then E is said to have locally finite perimeter. In that case the set function $U \mapsto P(E, U)$ extends to a positive Radon measure, denoted by

$$|D\chi_E|.$$

This is the *perimeter measure* of E.

If E has a smooth boundary, the divergence theorem gives

$$|D\chi_E| = \mathcal{H}^{n-1} \lfloor_{\partial E}.$$

For a general measurable set, however, perimeter measure can be completely different from Hausdorff measure on the topological boundary.

For example, let $B = B(0, 1)$ and set

$$E = B \setminus \mathbb{Q}^n.$$

Since E and B differ by a Lebesgue-null set,

$$\chi_E = \chi_B \qquad \text{almost everywhere.}$$

Their distributional derivatives, and hence their perimeter measures, are equal:

$$|D\chi_E| = |D\chi_B| = \mathcal{H}^{n-1} \lfloor_{\partial B}.$$

On the other hand, every point of $\overline{B}$ is a topological boundary point of E, so

$$\partial E = \overline{B}.$$

Consequently,

$$\mathcal{H}^{n-1}(\partial E) = \infty,$$

even though

$$|D\chi_E|(\mathbb{R}^n) = \mathcal{H}^{n-1}(\partial B) < \infty.$$

[10] H. Federer, *Curvature measures*, Trans. Amer. Math. Soc. **93** (1959), 418–491.

Perimeter measure depends only on the set up to a Lebesgue-null modification, whereas the topological boundary does not.[11]

4.4. **Three cautions and counterexamples.** The preceding gallery suggests several distinctions that are easy to miss if Lebesgue measure is one's only model.

Zero mass at every point does not determine the measure. Both Lebesgue measure and Cantor measure assign measure zero to every singleton, but they are mutually singular. More drastically, let X be uncountable and let

$$\mathcal{C} := \{A \subseteq X : A \text{ is countable or } A^c \text{ is countable}\}.$$

Then $\mathcal{C}$ is a σ-algebra, and

$$\mu(A) = \begin{cases} 0, & A \text{ is countable}, \\ 1, & A^c \text{ is countable} \end{cases}$$

defines a probability measure. Every singleton has measure zero although $\mu(X) = 1$. There is no contradiction: countable additivity says nothing about an uncountable union of null sets.

A Borel measure need not be complete. Lebesgue measure restricted to the Borel σ-algebra is not complete. Indeed, the Cantor set is a null Borel set, has cardinality $\mathfrak{c}$, and therefore has $2^{\mathfrak{c}}$ subsets, whereas $\mathbb{R}$ has only $\mathfrak{c}$ Borel sets. Some subsets of the Cantor set are consequently not Borel. Passing to the completion adds every subset of every Borel null set. Thus "Borel measurable" and "Lebesgue measurable" are not synonymous, even on $\mathbb{R}$.

Natural density is not a measure. For certain $A \subseteq \mathbb{N}$, the limit

$$d(A) := \lim_{N \to \infty} \frac{|A \cap \{1, \dots, N\}|}{N}$$

exists and is useful in number theory. The collection of sets for which $d(A)$ exists is not a σ-algebra, so natural density is not a measure even before countable additivity is considered. Moreover, no countably additive probability measure on $\mathcal{P}(\mathbb{N})$ can agree with natural density: every singleton would have measure zero while $\mathbb{N}$ would have measure one. Finitely additive extensions of density exist, but they necessarily give up one of the central axioms of the present theory.

5. Signed and complex measures

If $\mathcal{M}$ is a σ-algebra of subsets of X, a signed measure ν on $\mathcal{M}$ is a function

$$\nu : \mathcal{M} \to [-\infty, \infty],$$

allowed to assume at most one of the values $\pm\infty$ and satisfying $\nu(\emptyset) = 0$ and

$$S_j \in \mathcal{M} \text{ disjoint sequence} \implies \nu\left(\bigcup_j S_j\right) = \sum_j \nu(S_j),$$

the series $\nu(S_j)$ being absolutely convergent. Till now we have dealt with "positive measures". If μ_1, μ_2 are positive measures and one of them is finite, then $\mu_1 - \mu_2$ is a signed measure, for example.

If ν is a signed measure on $(X, \mathcal{M})$ and $E \in \mathcal{M}$, we say that ν is E-positive (negative) if $\nu(F) \geq 0$ (respectively, $\nu(F) \leq 0$) for $F \in \mathcal{M}, F \subseteq E$. We first prove the following

Lemma 5.1. *Suppose that ν does not take on the value ∞. If $\nu(A) > -\infty$, then there exists a measurable $P \subseteq A$ such that ν is P-positive and $\nu(P) \geq \nu(A)$.*

[11] See E. De Giorgi, *Su una teoria generale della misura $(r-1)$-dimensionale in uno spazio ad r dimensioni*, Ann. Mat. Pura Appl. **36** (1954), 191–213.

Proof. We start by proving the following: if $\nu(A) > -\infty$, then for all $\varepsilon > 0$, there exist measurable $B \subseteq A$ such that $\nu(B) \geq \nu(A)$ and $\nu(E) \geq -\varepsilon$ for all measurable $E \subseteq B$. Suppose this is false. Then, taking $B = A$, we see that for some $\varepsilon > 0$, there is a measurable $E_1 \subseteq A$ such that $\nu(E_1) \leq -\varepsilon$. Thus, $\nu(A \setminus E_1) \geq \nu(A) + \varepsilon$. Next, considering $B = A \setminus E_1$, we have a measurable $E_2 \subseteq A \setminus E_1$ such that $\nu(E_2) \leq -\varepsilon$, so $\nu(A \setminus (E_1 \cup E_2)) \geq \nu(A) + 2\varepsilon$. Continuing this way, we have $A_j := A \setminus (E_1 \cup \cdots \cup E_j) \searrow F, \nu(F) = \infty$. This gives a contradiction.

Now, define a sequence of sets $A_j \in \mathcal{M}$ inductively. Let $A_1 = A$. Given $A_j, j \leq n-1$, take $A_n \subseteq A_{n-1}$ such that $\nu(A_n) \geq \nu(A_{n-1})$ and $\nu(E) \geq -1/n$, for all measurable $E \subseteq A_n$. Then let $P = \bigcap A_j$. Clearly ν is P-positive, and $\nu(P) = \lim v(A_j) \geq \nu(A)$[12]. □

Now we prove the important

Theorem 5.2 (Hahn decomposition)**.** *If ν is a signed measure on $(X, \mathcal{M})$ then there exist $P, N \in \mathcal{M}$ such that ν is P-positive and N-negative and $P \cup N = X$, $P \cap N = \emptyset$.*

Proof. Given the above lemma, we proceed as follows. Assume without loss of generality that $s = \sup\{\nu(A) : A \in \mathcal{M}\} \geq 0$, otherwise there is nothing to prove. Take $A_j \in \mathcal{M}$ such that $\nu(A_j) \nearrow s, \nu(A_j) > -\infty$. By the lemma, ν is P-positive on a sequence of sets $P_j \subseteq A_j$ such that $\nu(P_j) \to s$. Set $P = \bigcup P_j$. One can easily check that ν is P-positive. Then we see that $\nu(P) = s < \infty$. Set $N = X \setminus P$. If N is not ν-negative, then there is a measurable set $S \subseteq N$ with $\nu(S) > 0$, and then $\nu(P \cup S) = s + \nu(S) > s$, which is not possible. □

If ν is a signed measure on $(X, \mathcal{M})$ and we have a Hahn partition $X = P \cup N$, define

$$\nu_+(E) = \nu(P \cap E), \quad \nu_-(E) = -\nu(N \cap E), E \in \mathcal{M}. \tag{44}$$

Since ν_+ and ν_- have disjoint supports, this is called the Hahn decomposition of the measure ν.

Exercise 5.3. *Prove that the Hahn decomposition is unique.*

Also, define

$$|\nu| = \nu_+ + \nu_-, \tag{45}$$

called the total variation measure of ν.

Given a signed measure ν, we can define $\int f \, d\nu$ for a suitable class of measurable functions f. We can use the Hahn decomposition to do this, setting

$$\int f \, d\nu = \int f \, d\nu_+ - \int f \, d\nu_-. \tag{46}$$

In particular, this is well-defined for integrable f.

One can also consider complex-valued measures on $(X, \mathcal{M})$, which is defined to be a function

$$\rho : \mathcal{M} \to \mathbb{C}, \tag{47}$$

satisfying

$$\rho(\emptyset) = 0 \tag{48}$$

and

$$E_j \in \mathcal{M} \text{ disjoint sequence } \implies \rho\left(\bigcup_j E_j\right) = \sum_j \rho(E_j). \tag{49}$$

In this case, ρ does not assume any infinite values. Clearly, one can decompose into real valued measures and define integrals in the obvious way.

[12]Here we have used the fact that if ν is a signed measure on $(X, \mathcal{M})$, and $E_j \in \mathcal{M}$, then $E_j \searrow E$, $\nu(E_1)$ finite $\implies \nu(E) = \lim_{j\to\infty} \nu(E_j)$. We leave this as an exercise.

6. L^p spaces

Let (X, μ) be a measure space. For $p \in [1, \infty)$, we define $L^P(X, \mu)$ to consist of measurable functions f such that

$$\left(\int_X |f(x)|^p \, d\mu\right)^{1/p} < \infty. \tag{50}$$

We define the left hand side to be the L^p-norm of f. We shall show that L^p are Banach spaces, but observe that technically, $L^p(X)$ consists of equivalence classes of functions, as $\|f - g\|_{L^p} = 0$ if (and only if) $f = g$ a.e. To verify that L^p-norms are indeed norms, the only non-trivial part is to check that it satisfies the triangle inequality, which we leave below as an exercise.

Exercise 6.1 (Triangle inequality)**.** *Prove that if $f, g \in L^p(X, \mu)$, then*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}. \tag{51}$$

Proof. Since $\|\alpha f\|_{L^p} = |\alpha| \|f\|_{L^p}$, by normalising the right hand side we can assume without loss of generality that $\|f\|_{L^p} = 1 - \lambda, \|g\|_{L^p} = \lambda$. Let $\tilde{f} := \frac{f}{1-\lambda}$ and $\tilde{g} = \frac{g}{\lambda}$. Then, it suffices to prove that

$$\|(1 - \lambda)\tilde{f} + \lambda\tilde{g}\|_{L^p}^p \leq 1$$

when $\|\tilde{f}\|_{L^p} = \|\tilde{g}\|_{L^p} = 1$. Since $x \mapsto |x|^p$ is convex, it follows immediately that $\left|(1 - \lambda)\tilde{f} + \lambda\tilde{g}\right|^p \leq (1 - \lambda)|\tilde{f}|^p + \lambda|\tilde{g}|^p$. This proves our claim[13].

□

The L^p spaces give interesting examples of Banach spaces. Below, we give a sketch of the proof.

Theorem 6.2. *For $p \in [1, \infty)$, $L^p(X, \mu)$ with the norm (50), is a Banach space.*

Proof. Consider a Cauchy sequence $f_n \in L^p(X, \mu)$, and assume (passing to a subsequence if need be) that $\|f_{n+1} - f_n\|_{L^p} \leq 2^{-n}$. Let $G_m(x) = \sum_{n=1}^m |f_{n+1}(x) - f_n(x)|$, and $G = \sum_n^\infty |f_{n+1}(x) - f_n(x)|$.

$G_m \nearrow G$ (maybe G is infinity-valued, a priori), and in L^p-norm. This is because, $|G - G_m|^p \searrow 0$ is a sequence of non-negative functions[14]. Since each G_m satisfies $\|G_m\|_{L^p} \leq 1$ (follows from triangle inequality), we have that $\int |G|^p \leq 1$, which means that $G < \infty$ a.e.. Hence the series

$$f_1(x) + \sum_{n=1}^{k} (f_{n+1}(x) - f_n(x)) = f_{k+1}(x) \to f(x) \text{ a.e.}$$

Since $|f - f_{m+1}| \leq G - G_m$, DCT gives us that

$$\int_X |f - f_{m+1}|^p \leq \int_X (G - G_m)^p \to 0, \text{ as } m \to \infty.$$

□

The following density theorems are highly useful.

Exercise 6.3 (Density of simple and continuous functions)**.** *Let $(X, \mathcal{M}, \mu)$ be a positive measure space and let $1 \leq p < \infty$.*

(1) *Prove that the simple functions s satisfying*

$$\mu\big(\{x \in X : s(x) \neq 0\}\big) < \infty$$

are dense in $L^p(X, \mu)$.

[13] I found this slick proof from Tao's blog: https://terrytao.wordpress.com/2009/01/09/245b-notes-3-lp-spaces/.

[14] We are using the version of the MCT which says that if f_n is a sequence of non-negative decreasing measurable functions and $\int f_1 < \infty$, then limits and integrals are interchangeable. This follows from the standard MCT by considering the sequence $f_1 - f_n$. The condition $\int f_1 < \infty$ is required. Otherwise, the sequence $f_n = \chi_{[n,\infty]}$ on $\mathbb{R}$ would give a counterexample.

(2) *Now suppose that X is a locally compact Hausdorff space and that μ is a Radon Borel measure. Prove that $C_c(X)$ is dense in $L^p(X,\mu)$.*

Hint for (1). *For a non-negative $f \in L^p(X,\mu)$, use the simple approximations φ_j from Remark 1.27. They satisfy $0 \le \varphi_j \le f$, $\varphi_j \to f$ pointwise, and*

$$\{\varphi_j \neq 0\} \subseteq \{f \ge 2^{-j}\}.$$

Chebyshev's inequality gives

$$\mu\big(\{f \ge 2^{-j}\}\big) \le 2^{jp} \int_X |f|^p \, d\mu < \infty.$$

The Dominated Convergence Theorem then gives $\varphi_j \to f$ in L^p. Apply this separately to the positive and negative parts, or to the real and imaginary parts.

Hint for (2). *It is enough, by part (1), to approximate a simple function whose support has finite measure. First consider a measurable set E with $\mu(E) < \infty$. By regularity, choose a compact set $K \subseteq E$ and a relatively compact open set $V \supseteq K$ such that*

$$\mu(E \setminus K) + \mu(V \setminus K)$$

is arbitrarily small. By Urysohn's lemma, choose $\phi \in C_c(X)$ with

$$0 \le \phi \le 1, \qquad \phi = 1 \text{ on } K, \qquad \operatorname{supp}\phi \subseteq V.$$

Then

$$\|\chi_E - \phi\|_{L^p}^p \le \mu(E \setminus K) + \mu(V \setminus K).$$

Thus characteristic functions of finite-measure sets can be approximated by functions in $C_c(X)$. Approximate each term of a finite-support simple function in this way.

The case of $p = 2$ is particularly important, as L^2 spaces have an extra structure, namely, that of a Hilbert space.

Definition 6.4. *A Hilbert space H is a complete inner product space. Recall that an inner product is a bilinear form $H \times H \mapsto \mathbb{K}$ such that $(u,v) = \overline{(v,u)}$ and $(v,v) = 0 \iff v = 0$.*

A Hilbert space trivially becomes a Banach space with the norm $\|v\| = (v,v)^{1/2}$. Morally, a Hilbert space is a Banach space with the added concept of "angles", which allows us to talk about orthogonal complements.

The archetypal example of a (complex) Hilbert space is $L^2(X,\mu)$, where μ is any positive measure. The corresponding inner product is given by

$$(f,g) = \int_X f\overline{g} \, d\mu.$$

Exercise 6.5. *Show that the functions $e^{in\theta}, n \in \mathbb{Z}$ form a complete ONB of $L^2(S^1)$.*

Exercise 6.6. *Consider the open unit disc $\mathbb{D} \subset \mathbb{C}$. Check that the polynomials $z^n, n \in \mathbb{N} \cup \{0\}$ are orthogonal in $L^2(\mathbb{D})$. After scaling them, can you make them a Schauder basis? If not, what more is required?*

We now prove the Riesz representation theorem, which is astoundingly simple, yet one of the most useful results in analysis.

Theorem 6.7 (Riesz representation theorem)**.** *If $T : H \to \mathbb{C}$ is a continuous linear map, there exists a unique $v \in H$ such that $T(u) = (u,v)$ for all $u \in H$.*

Proof. Consider $K = \operatorname{Ker} T$, and convince yourself that it is a closed linear subspace of H. WLOG, $K^\perp \neq 0$. Select $x \in K^\perp$ so that $Tx = 1$. For all $y \in K^\perp$, $y - (Ty)x \in K \cap K^\perp$, so $K^\perp$ is one-dimensional (spanned by x). Finish the argument. What multiple of x should we set as v? (Hint: If $v = \alpha x$, then $\alpha = 1/\|x\|^2$). □

Exercise 6.8. *(optional for $n \geq 2$): Let $\mathbb{D}_n$ denote the open unit ball in $\mathbb{C}^n$ and $\mathcal{A}(\mathbb{D}_n)$ denote the space of complex analytic functions. Let $H := \mathcal{A}(\mathbb{D}_n) \cap L^2(\mathbb{D}_n)$.*
(a) Prove that H is a Hilbert space with the inner product restricted from $L^2(\mathbb{D}_n)$.
(b) Prove that for any $p \in \mathbb{D}_n$, point evaluation $L_p : H \to \mathbb{C}$, given by $L_p(f) = f(p)$ is a continuous linear functional.
(c) By an application of the Riesz representation theorem, L_p is given by inner product with an element $K_p \in H$, that is,

$$f(p) = \int_{\mathbb{D}_n} K_p \overline{f}.$$

Evaluate $K_p(z)$. This is known as the Bergman kernel. (This foregoing construction is part of a larger theory of "reproducing kernel Hilbert spaces").

As a quick application of the Riesz representation theorem, we prove the celebrated Radon-Nikodym theorem. Let ν and μ be nonnegative measures on a σ-algebra on a measure space X. ν is said to be *absolutely continuous* with respect to μ, denoted by $\nu \ll \mu$, if every set that has μ-measure zero also has ν-measure zero.

Theorem 6.9 (Radon-Nikodym theorem)**.** *In the above setting, we have a non-negative integrable function g such that*

$$\nu(E) = \int_E g \, d\mu. \tag{52}$$

Furthermore, if ν is additionally allowed to be a complex measure, then (52) still holds, but now we can only say that $g \in L^1(X, \mu)$.

The proof below is due to von Neumann.

Proof. Assume for simplicity that μ and ν are finite measures. Let H be the real Hilbert space $L^2_{\mu+\nu}$. Define the linear functional $Tf = \int f \, d\mu$, which is $\leq C\|f\|_{L^2_\mu}$, which in turn is $\leq C\|f\|_{L^2_{\mu+\nu}}$. Then, by Theorem 6.7,

$$Tf = \int fh \, d(\mu + \nu), \quad h \in H. \tag{53}$$

Rewriting this as

$$\int f(1-h) \, d\mu = \int fh \, d\nu, \tag{54}$$

we wish to claim that $0 < h \leq 1$ except on a set of μ-measure zero. Let F denote the set on which $h \leq 0$. Set $f = 1$ on F, and $f = 0$ outside F. This gives from (53) that

$$\int_F d\mu = \int_F h \, d(\mu + \nu),$$

which implies that $h \leq 0$ on a set of μ-measure 0. Also, let G denote the set where $h > 1$. Set $f = 1$ on G and $f = 0$ outside G. By a similar calculation, it is clear that

$$0 < h \leq 1 \tag{55}$$

holds μ-everywhere. One can redefine h on a measure (with respect to μ and ν also) 0 such that (55) holds everywhere. Now, we choose $f = \chi_E/h$ in (53). This gives that

$$\int_E \frac{1}{h} \, d\mu = \int_E d(\mu + \nu),$$

which gives that

$$\nu(E) = \int_E (1/h - 1) \, d\mu. \tag{56}$$

So it is clear that we have to choose

$$g = \frac{1-h}{h},$$

and setting $E = X$ in (56), we see that $g \in L^1(X, \mu)$.

The complex measure case follows immediately by applying the above argument to the positive and negative variations of both real and imaginary parts of ν separately. □

Exercise 6.10. *Extend the Radon-Nikodym theorem to the setting of σ-finite measure spaces*[15]*. This is the same trick from [Rud87, Lemma 6.9, Chapter 6] that we discussed earlier while talking about spectral measures.*

Throughout, $B(x, r)$ denotes the open Euclidean ball of radius r centred at x, and $|E| := \mathcal{L}^d(E)$.

Lemma 6.11 (Finite covering lemma)**.** *Let $B_1, \dots, B_N$ be finitely many balls in $\mathbb{R}^d$. There is a pairwise disjoint subcollection $B_{i_1}, \dots, B_{i_m}$ such that*

$$\bigcup_{i=1}^{N} B_i \subseteq \bigcup_{j=1}^{m} 3B_{i_j},$$

where $3B$ denotes the ball with the same centre as B and three times its radius.

Proof. Relabel the balls so that their radii are nonincreasing. Retain B_1, and then retain B_i precisely when it is disjoint from all previously retained balls. The retained balls are pairwise disjoint. If B_i is not retained, then it intersects a previously retained ball B_j whose radius satisfies $r_j \geq r_i$. For every $y \in B_i$,

$$|y - x_j| \leq |y - x_i| + |x_i - x_j| < r_i + (r_i + r_j) \leq 3r_j.$$

Thus $B_i \subseteq 3B_j$, which proves the result. □

Lemma 6.12 (Hardy–Littlewood maximal inequality)**.** *For $f \in L^1(\mathbb{R}^d)$, define*

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

Then, for every $t > 0$,

$$\big|\{x \in \mathbb{R}^d : Mf(x) > t\}\big| \leq \frac{3^d}{t} \|f\|_{L^1(\mathbb{R}^d)}.$$

Proof. For each fixed $r > 0$, the function

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

is continuous, since it is the convolution of $|f|$ with the normalized characteristic function of $B(0, r)$. Hence Mf, being the supremum of continuous functions, is lower semicontinuous, and

$$E_t := \{x \in \mathbb{R}^d : Mf(x) > t\}$$

is open.

Let $K \subseteq E_t$ be compact. For every $x \in K$, choose $r_x > 0$ such that

$$\int_{B(x,r_x)} |f(y)|\, dy > t|B(x, r_x)|.$$

[15] As far as I can see, one cannot get integrable g here, only measurable. But I strongly encourage you to think about this, as I could be wrong!

Finitely many of these balls cover K. By the finite covering lemma, there is a pairwise disjoint subcollection $B_1, \dots, B_m$ such that

$$K \subseteq \bigcup_{j=1}^{m} 3B_j.$$

Therefore

$$|K| \le \sum_{j=1}^{m} |3B_j| = 3^d \sum_{j=1}^{m} |B_j| < \frac{3^d}{t} \sum_{j=1}^{m} \int_{B_j} |f(y)|\, dy \le \frac{3^d}{t} \|f\|_{L^1(\mathbb{R}^d)}.$$

Since E_t is open, inner regularity of Lebesgue measure gives

$$|E_t| = \sup\{|K| : K \subseteq E_t,\ K \text{ compact}\} \le \frac{3^d}{t} \|f\|_{L^1(\mathbb{R}^d)}.$$

□

Theorem 6.13 (Lebesgue differentiation theorem)**.** *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$. Then, for almost every $x \in \mathbb{R}^d$,*

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

Consequently,

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

for almost every $x \in \mathbb{R}^d$.

Proof. Suppose first that $f \in L^1(\mathbb{R}^d)$, and define

$$Df(x) := \limsup_{r \downarrow 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)|\, dy$$

at every point where $f(x)$ is finite. For each such x, the expression inside the limsup is continuous as a function of $r > 0$. Hence the limsup may be taken over rational radii, and Df is measurable.

Fix $\varepsilon > 0$. By the density of $C_c(\mathbb{R}^d)$ in $L^1(\mathbb{R}^d)$, choose $\varphi \in C_c(\mathbb{R}^d)$ such that

$$\|f - \varphi\|_{L^1(\mathbb{R}^d)} < \varepsilon.$$

For almost every x and every $r > 0$,

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

Since φ is continuous,

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

It follows that

$$Df(x) \le M(f - \varphi)(x) + |f(x) - \varphi(x)|$$

for almost every x. Therefore, for every $\alpha > 0$,

$$\{Df > \alpha\} \subseteq \left\{M(f-\varphi) > \frac{\alpha}{2}\right\} \cup \left\{|f - \varphi| > \frac{\alpha}{2}\right\}$$

up to a null set. The Hardy–Littlewood maximal inequality and Chebyshev's inequality give

$$|\{Df > \alpha\}| \le \frac{2 \cdot 3^d}{\alpha} \|f - \varphi\|_{L^1} + \frac{2}{\alpha} \|f - \varphi\|_{L^1} < \frac{2(3^d + 1)}{\alpha} \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary,

$$|\{Df > \alpha\}| = 0$$

for every $\alpha > 0$. Hence

$$\{Df > 0\} = \bigcup_{m=1}^{\infty} \{Df > 1/m\}$$

has measure zero, and therefore $Df = 0$ almost everywhere.

Now let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$. For each $N \in \mathbb{N}$, set

$$f_N := f\chi_{B(0,N)}.$$

Then $f_N \in L^1(\mathbb{R}^d)$, so almost every point of $\mathbb{R}^d$ is a Lebesgue point of f_N. If $x \in B(0,N)$, then $f_N = f$ on every sufficiently small ball centred at x. Thus almost every point of $B(0,N)$ is a Lebesgue point of f. Since

$$\mathbb{R}^d = \bigcup_{N=1}^{\infty} B(0,N),$$

the first conclusion follows.

Finally,

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

which proves the second conclusion. □

Definition 6.14. *Let X be a topological space. A Borel measure μ on X is called* locally finite *if every $x \in X$ has an open neighbourhood U such that $\mu(U) < \infty$.*

If X is locally compact and Hausdorff, local finiteness is equivalent to

$$\mu(K) < \infty$$

for every compact set $K \subseteq X$. Indeed, if μ is locally finite, a compact set can be covered by finitely many open sets of finite measure. Conversely, if μ is finite on compact sets, local compactness gives, for every $x \in X$, an open neighbourhood U whose closure is compact, and hence

$$\mu(U) \le \mu(\overline{U}) < \infty.$$

In particular, a Borel measure μ on $\mathbb{R}^d$ is locally finite if and only if

$$\mu(B(0,R)) < \infty$$

for every $R > 0$. Such a measure is automatically σ-finite because

$$\mathbb{R}^d = \bigcup_{N=1}^{\infty} B(0,N).$$

Corollary 6.15. *Let μ be a locally finite positive Borel measure on $\mathbb{R}^d$ such that $\mu \ll \mathcal{L}^d$, and let*

$$g := \frac{d\mu}{d\mathcal{L}^d}.$$

Then

$$\lim_{r \downarrow 0} \frac{\mu(B(x,r))}{|B(x,r)|} = g(x)$$

for $\mathcal{L}^d$-almost every $x \in \mathbb{R}^d$.

Proof. Local finiteness of μ implies that $g \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$. Moreover,

$$\frac{\mu(B(x,r))}{|B(x,r)|} = \frac{1}{|B(x,r)|} \int_{B(x,r)} g(y)\,dy.$$

The result follows from the Lebesgue differentiation theorem. □

Remark 6.16. *The corollary concerns differentiation with respect to Lebesgue measure. The corresponding statement*

$$\lim_{r\downarrow 0}\frac{\mu(B(x,r))}{\nu(B(x,r))}=\frac{d\mu}{d\nu}(x)$$

for arbitrary locally finite Radon measures $\mu \ll \nu$ on $\mathbb{R}^d$ is the Besicovitch differentiation theorem.

Proposition 6.17 (L^p-L^q duality)**.** *Let (X,μ) be a finite measure space and let $1 \le p < \infty$. If q is defined by*

$$\frac{1}{p}+\frac{1}{q}=1,$$

with $q=\infty$ when $p=1$, then $L^p(X,\mu)^*$ is naturally isometrically isomorphic to $L^q(X,\mu)$.

Proof. For $v \in L^q(X,\mu)$, define

$$T_v : L^p(X,\mu) \longrightarrow \mathbb{C}, \qquad T_v u := \int_X uv\, d\mu.$$

By Hölder's inequality,

$$|T_v u| \le \|u\|_{L^p}\|v\|_{L^q},$$

and hence $\|T_v\| \le \|v\|_{L^q}$.

We first prove the reverse inequality. Suppose that $1 < p < \infty$ and $v \neq 0$. Define

$$\alpha(x) := \begin{cases} \dfrac{\overline{v(x)}}{|v(x)|}, & v(x)\neq 0, \\ 0, & v(x)=0, \end{cases} \qquad u := \frac{\alpha|v|^{q-1}}{\|v\|_{L^q}^{q-1}}.$$

Since $(q-1)p = q$, one has $\|u\|_{L^p}=1$, while

$$T_v u = \frac{1}{\|v\|_{L^q}^{q-1}}\int_X |v|^q\, d\mu = \|v\|_{L^q}.$$

Thus $\|T_v\| \ge \|v\|_{L^q}$.

Now suppose that $p=1$. If $v\neq 0$, then for every $0<\varepsilon<\|v\|_{L^\infty}$, the set

$$E_\varepsilon := \{x \in X : |v(x)| > \|v\|_{L^\infty}-\varepsilon\}$$

has positive measure. With α defined as above, let

$$u_\varepsilon := \frac{\alpha\chi_{E_\varepsilon}}{\mu(E_\varepsilon)}.$$

Then $\|u_\varepsilon\|_{L^1}=1$ and

$$|T_v u_\varepsilon| = \frac{1}{\mu(E_\varepsilon)}\int_{E_\varepsilon}|v|\, d\mu > \|v\|_{L^\infty}-\varepsilon.$$

Letting $\varepsilon \downarrow 0$ gives $\|T_v\| \ge \|v\|_{L^\infty}$. Therefore

$$\|T_v\| = \|v\|_{L^q}$$

in all cases, and the map $v \mapsto T_v$ is a linear isometry from $L^q(X,\mu)$ into $L^p(X,\mu)^*$.

It remains to prove surjectivity. Let $T \in L^p(X,\mu)^*$ and define

$$\nu(E) := T(\chi_E)$$

for every measurable set $E \subseteq X$. This is well defined because $\mu(X)<\infty$, so $\chi_E \in L^p(X,\mu)$.

We verify countable additivity. Let $(E_j)_{j\ge 1}$ be pairwise disjoint measurable sets and set $E := \bigcup_{j=1}^\infty E_j$. Since

$$\left\|\chi_E - \sum_{j=1}^N \chi_{E_j}\right\|_{L^p}^p = \mu\left(E \setminus \bigcup_{j=1}^N E_j\right) \longrightarrow 0,$$

continuity of T gives

$$\nu(E) = \lim_{N\to\infty} T\left(\sum_{j=1}^{N} \chi_{E_j}\right) = \lim_{N\to\infty} \sum_{j=1}^{N} \nu(E_j).$$

Thus ν is a complex measure. Moreover,

$$|\nu(E)| = |T(\chi_E)| \le \|T\|\|\chi_E\|_{L^p} = \|T\|\mu(E)^{1/p}.$$

In particular, $\nu \ll \mu$. By the Radon–Nikodym theorem for complex measures, there exists $w \in L^1(X,\mu)$ such that

$$\nu(E) = \int_E w\, d\mu$$

for every measurable set E. Consequently,

$$Tf = \int_X f\, d\nu = \int_X fw\, d\mu$$

for every simple function f.

This identity also holds for every $f \in L^\infty(X,\mu)$. Indeed, choose simple functions f_j converging uniformly to f. Since $\mu(X) < \infty$,

$$\|f_j - f\|_{L^p} \le \mu(X)^{1/p}\|f_j - f\|_{L^\infty} \longrightarrow 0,$$

and, since $w \in L^1(X,\mu)$,

$$\left|\int_X (f_j - f)w\, d\mu\right| \le \|f_j - f\|_{L^\infty}\|w\|_{L^1} \longrightarrow 0.$$

Passing to the limit gives

$$Tf = \int_X fw\, d\mu$$

for every $f \in L^\infty(X,\mu)$.

Suppose first that $p = 1$. Define

$$\alpha(x) := \begin{cases} \dfrac{\overline{w(x)}}{|w(x)|}, & w(x) \ne 0, \\ 0, & w(x) = 0. \end{cases}$$

For $a > \|T\|$, let $E_a := \{x \in X : |w(x)| > a\}$. If $\mu(E_a) > 0$, then $f := \alpha\chi_{E_a}$ belongs to $L^\infty(X,\mu)$ and

$$a\mu(E_a) < \int_{E_a} |w|\, d\mu = Tf \le |Tf| \le \|T\|\|f\|_{L^1} = \|T\|\mu(E_a),$$

which is impossible. Hence $\mu(E_a) = 0$ for every $a > \|T\|$, and therefore

$$\|w\|_{L^\infty} \le \|T\|.$$

Now suppose that $1 < p < \infty$. Define α as above and, for $n \in \mathbb{N}$, let

$$E_n := \{x \in X : |w(x)| \le n\}, \qquad f_n := \chi_{E_n}|w|^{q-1}\alpha.$$

Then $f_n \in L^\infty(X,\mu)$ and

$$|f_n|^p = \chi_{E_n}|w|^q.$$

Thus, writing

$$A_n := \int_{E_n} |w|^q\, d\mu,$$

we obtain

$$A_n = \int_X f_n w \, d\mu = |Tf_n| \leq \|T\| \|f_n\|_{L^p} = \|T\| A_n^{1/p}.$$

If $A_n > 0$, it follows that $A_n^{1/q} \leq \|T\|$, and the same conclusion is trivial when $A_n = 0$. Since $E_n \uparrow X$ up to a null set, the Monotone Convergence Theorem gives

$$\|w\|_{L^q}^q = \int_X |w|^q \, d\mu = \lim_{n\to\infty} A_n \leq \|T\|^q.$$

Hence $\|w\|_{L^q} \leq \|T\|$.

In either case, $w \in L^q(X, \mu)$. Both maps

$$f \longmapsto Tf \qquad \text{and} \qquad f \longmapsto \int_X fw \, d\mu$$

are continuous on $L^p(X, \mu)$ and agree on the simple functions, which are dense in $L^p(X, \mu)$. Therefore

$$Tf = \int_X fw \, d\mu$$

for every $f \in L^p(X, \mu)$. Thus $T = T_w$, proving surjectivity. Finally, since $v \mapsto T_v$ is an isometry,

$$\|T\| = \|T_w\| = \|w\|_{L^q}.$$

□

Question: Which step in the above proof fails for $p = \infty$? (Hint: pay particular attention to the countable additivity claim above).

Exercise 6.18. *Extend the above result to σ-finite measures μ.*

Since the L^p spaces have Banach structure, it is of importance to try and find bounded linear operators. A useful class of such operators is given in practice by integral operators.

Proposition 6.19 (Boundedness of integral operators)**.** *Let*

$$Kf(x) = \int_X k(x, y) f(y) \, d\mu(y),$$

where (X, μ) is a measure space. K defines a bounded operator on $L^p(X)$ for $p \in [1, \infty]$ if

$$k \text{ is measurable on } X \times X,$$

and

$$\int_X |k(x, y)| dx \leq C_1, \quad \int_X |k(x, y)| dy \leq C_2.$$

Proof. We are only going to cover the cases $p \in (1, \infty)$. Let q be such that $1/p + 1/q = 1$. Let $\|f\|_{L^p} = 1$. Then

$$\int_X |k(x, y) f(y)| \, dy = \int_X |k(x, y)|^{1/q} |k(x, y)|^{1/p} |f(y)| \, dy \leq C_2^{1/q} \left(\int_X |k(x, y)| |f(y)|^p \, dy \right)^{1/p},$$

which implies that

$$\begin{aligned} \|Kf\|_{L^p} &\leq C_2^{1/q} \left(\int_X \int_X |k(x, y)| |f(y)|^p \, dy dx \right)^{1/p} \\ &= C_2^{1/q} \left(\int_X |f(y)|^p \left(\int_X |k(x, y)| \, dx \right) \, dy \right)^{1/p} \\ &\leq C_2^{1/q} C_1^{1/p}. \end{aligned}$$

□

These are called integral operators, and $k(x, y)$ is the integral kernel.

Exercise 6.20. *Address the cases $p = 1, \infty$.*

Exercise 6.21. *In the Hilbert space $L^2(X, \mu)$, K^* is an integral operator with kernel $k^*(x, y) = \overline{k(y, x)}$.*

Now we define a more general linear topological space, which is general enough to meet many practical examples, but still nice enough to allow most theorems to go through.

Definition 6.22 (Fréchet spaces)**.** *Take a vector space V and a countable family of seminorms $p_j : V \to \mathbb{R}^+$[16], where a seminorm satisfies all the properties of a norm with the exception of $\|v\| = 0 \implies v = 0$. We impose the additional restriction that given a non-zero $x \in V$, there exists some semi-norm p_k such that $p_k(x) \neq 0$[17]. Now, setting*

$$d(u, v) = \sum_{j=0}^{\infty} 2^{-j} \frac{p_j(u - v)}{1 + p_j(u - v)} \tag{57}$$

we have a metric on V[18]. If V is a complete metric space with respect to this structure, then V is a Fréchet space.

Remark 6.23. *If $B \in \mathcal{B}$ is a local open base at $0 \in \mathbb{K}$, then sets $p_j^{-1}(B), B \in \mathcal{B}, j \in \mathbb{N} \cup \{0\}$ will generate a subbase at the origin of V. This can also be used to define a topology on V. You should check that this is the same as the topology generated by the metric in (57).*

Examples: 1. The paradigmatic example of a Fréchet space is that of smooth functions on Riemannian manifolds, with or without boundary. Here, the seminorms p_j are given by $\|.\|_{C^j(M)}$. On compact Riemannian manifolds, these seminorms are actually norms, but one gets legitimate seminorms in the case of non-compact manifolds M. Express $M = \cup_k M_k$, where M_k is an increasing sequence of compact manifolds with boundary, and consider the seminorms $p_j = \|.\|_{C^j(M_j)}$. A familiar setting is $M = \mathbb{R}^n$ with $M_k = B_k(0)$ giving a compact exhaustion.
2. Similar ideas as above make $L^p_{\text{loc}}(M)$ Fréchet spaces for $p \in [1, \infty]$.
3. Let $\mathcal{A}(U)$ be the space of holomorphic functions on a domain $U \subseteq \mathbb{C}$. Then the seminorms $p_j(f) = \sup\{|f(z)| : |z| \leq j\}$ convert $\mathcal{A}(U)$ into a Fréchet space.
4. Another important example is the Schwartz space of rapidly decreasing functions

$$\mathcal{S}(\mathbb{R}^n) = \{u \in C^\infty(\mathbb{R}^n) : |D^\alpha u(x)| \leq C(1 + |x|)^{-N} \text{ for all } \alpha, N\},$$

with seminorms

$$p_k(u) = \sup_{x \in \mathbb{R}^n, |\alpha| \leq k} (1 + |x|^2)^{k/2} |D^\alpha u(x)|.$$

Exercise 6.24. *In a Fréchet space V, a set S is called* bounded *if each p_j is bounded on S. Show that every bounded subset of the Schwartz space $\mathcal{S}(\mathbb{R}^n)$ is relatively compact*[19]*.*

Exercise 6.25. *Prove that $\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ for all $p \in [1, \infty]$, and the inclusion is continuous. Prove further that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$. (Hint: one way to do this is to use the fact that $C_c(\mathbb{R}^n)$ is dense in L^p, and approximate a $C_c(\mathbb{R}^n)$ function with a function in $C_c^\infty(\mathbb{R}^n)$. To accomplish the latter, take convolution with a bump function of L^1-norm 1. Smoothness is immediate because of the well-known behaviour of differential operators with convolutions. Also, support of the convolution is contained inside the sum of the supports, so compact support follows).*

[16] Henceforth, $\mathbb{R}^+$ will denote the set of non-negative real numbers.

[17] Otherwise, there is a concern that the metric (57) below might become a pseudometric unless this is imposed.

[18] Check this!

[19] This is pretty interesting, as no infinite dimensional Banach space can have this property.

The Schwartz space is one of the most (arguably, the most) important and useful example of Fréchet spaces, and one fundamentally important aspect of the Schwartz space is the fact that it behaves very nicely with the Fourier transform. We use the following convention for the Fourier transform:

$$\hat{f}(\xi) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} f(x) e^{-ix.\xi}\, dx. \tag{58}$$

A priori, the Fourier transform is defined as a bounded operator $L^1 \mapsto L^\infty$. But we have the following:

Proposition 6.26 (Plancherel theorem)**.** *The Fourier transform $\mathcal{F} : S(\mathbb{R}^n) \to S(\mathbb{R}^n)$ is a linear isomorphism, which extends by continuity to a Hilbert space isomorphism of $L^2(\mathbb{R}^n)$*[20]*.*

Proof. Denoting by $(.,.)$ the usual L^2 inner product, it is clear that $\mathcal{F}^*$, defined by

$$\mathcal{F}^*(u)(x) = \frac{1}{(2\pi)^{n/2}} \int u(\xi) e^{ix.\xi}\, d\xi \quad \forall u \in \mathcal{S}(\mathbb{R}^n), \tag{59}$$

satisfies that

$$(\mathcal{F}(u), v) = (u, \mathcal{F}^*(v)) \quad \forall u, v \in \mathcal{S}(\mathbb{R}^n). \tag{60}$$

We will first show that $\mathcal{F}$ is bijective. By definition,

$$\begin{aligned}
\mathcal{F}^*\mathcal{F}f(x) &= \frac{1}{(2\pi)^n} \int\int f(y) e^{i(x-y).\xi}\, dy d\xi \\
&= \frac{1}{(2\pi)^n} \lim_{t \searrow 0} \int\int f(y)\, e^{-t|\xi|^2}\, e^{i(x-y).\xi}\, dy d\xi \quad \text{(applying DCT to the } y \text{ integral)} \\
&= \frac{1}{(2\pi)^n} \lim_{t \searrow 0} \int f(y) \int \left(e^{-t|\xi|^2}\, e^{i(x-y).\xi} \right) d\xi dy \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \lim_{t \searrow 0} \int f(y)\, \mathcal{F}^*(e^{-t|\xi|^2})(x-y)\, dy \\
&= \lim_{t \searrow 0} \frac{1}{(4\pi t)^{\frac{n}{2}}} \int f(y)\, e^{-|x-y|^2/4t}\, dy = f(x),
\end{aligned} \tag{61}$$

which proves the bijectivity of $\mathcal{F}$[21]. Now, it is clear that for all $u \in \mathcal{S}(\mathbb{R}^n)$,

$$\begin{aligned}
(u, u) &= (\mathcal{F}^*\mathcal{F}u, u) \\
&= (\mathcal{F}u, \mathcal{F}u) \quad \text{using (60)},
\end{aligned}$$

which basically proves that the Fourier transform is an isometric isomorphism on $\mathcal{S}(\mathbb{R}^n)$ with respect to the L^2 norm. Since $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$, we are done. □

Remark 6.27. *1. Observe that since the inner product on a Hilbert space can be recovered from the norm*[22] *it is clear that the Fourier transform is actually a unitary operator on $L^2(\mathbb{R}^n)$.*
2. Observe that $\mathcal{F}, \mathcal{F}^*$ do not provide inversion for the Fourier transform $L^1 \to L^\infty$. This will be taken care of later as a bi-product of the Fourier inversion formula for tempered distributions.

[20] Another way to achieve this is by defining the Fourier transform $\mathcal{F} : L^1 \to L^\infty$, and then extending by hand from $L^1 \cap L^2$ to L^2 via density arguments. For details, one can refer to Stein-Weiss. But I personally prefer the above approach, as it takes care of the isomorphism on tempered distributions immediately via duality arguments.

[21] For a justification of the last step, see the Appendix on page 9 of my functional analysis notes. Roughly, this basically means that the solution to the heat equation converges pointwise to the initial condition.

[22] Via the "polarization identity"

$$4(u, v) = \|u + v\|^2 - \|u - v\|^2.$$

6.1. **A few remarks around the heat equation.** Here we collect some facts that are going to become relevant later when we start doing probability.

Let us consider the heat equation $(\partial_t - \Delta)u(t,x) = 0$, with initial condition $u(0,x) = f(x)$. Fourier transforming in the x variable only, one gets $(\partial_t + |\xi|^2)\hat{u}(t,\xi) = 0$ with initial condition $u(0,\xi) = \hat{f}(\xi)$. Solving this as an ODE in the t variable, we get that $\hat{u}(t,\xi) = \hat{f}(\xi)e^{-t|\xi|^2}$. Again, using the well-known properties of Fourier transform, this reduces to

$$u(t,x) = f * \frac{1}{(4\pi t)^{n/2}} e^{-|.|^2/4t},$$

which is the expression on the left hand side of (61). The expression

$$p(t,x,y) := \frac{1}{(4\pi t)^{n/2}} e^{-|x-y|^2/4t} \tag{62}$$

is called the "heat kernel", and is one of the most central constructions in all of Mathematics[23].

(61) claims that $f * \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t} \to f(x)$ as $t \searrow 0$, or equivalently, $u(t,x) \to f(x)$ as $t \searrow 0$.

Proposition 6.28. *Let $u(t,x) = \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-\frac{|x-y|^2}{4t}} f(y)dy$ represent a solution of the heat equation $(\partial_t - \Delta)u = 0, u(0,x) = f(x)$. If f is continuous and bounded, then*

$$\lim_{t \searrow 0} u(t,x) = f(x).$$

Proof. We begin the quoting the well-known fact that $p(t,x,y) := \frac{1}{(4\pi t)^{n/2}} e^{-\frac{|x-y|^2}{4t}}$ is a probability density measure, that is[24],

$$\int_{\mathbb{R}^n} p(t,x,y)dy = 1.$$

Using this, now we have

$$\begin{aligned}
|f(x) - u(t,x)| &= \left| \int_{\mathbb{R}^n} (f(x) - f(y))p(t,x,y)dy \right| \\
&\leq \frac{1}{\pi^{n/2}} \int_{r=0}^{\infty} \int_{S^{n-1}} |f(x) - f(x - \sqrt{t}r\omega)| e^{-r^2} r^{n-1} dr d\omega \text{ (change variables } \frac{x-y}{\sqrt{t}} = r\omega) \\
&= \frac{1}{\pi^{n/2}} \int_{r=0}^{K} \int_{S^{n-1}} |f(x) - f(x - \sqrt{t}r\omega)| e^{-r^2} r^{n-1} dr d\omega + \\
&\frac{1}{\pi^{n/2}} \int_{r>K} \int_{S^{n-1}} |f(x) - f(x - \sqrt{t}r\omega)| e^{-r^2} r^{n-1} dr d\omega.
\end{aligned}$$

First choose large enough K and then t small enough. The first integral is made small by using the continuity of f, and the second integral is made small by the boundedness of f and the exponential decay of the function e^{-r^2}. □

Now note further that even if we start with f in $L^2(\mathbb{R}^n)$, we have the following:

Exercise 6.29. *Prove that $u(t,x)$ is smooth in x. Furthermore, prove that $u(t,x)$ is real analytic for x and $t > 0$.*

Here is a slick proof of the above exercise:

[23] In particular, it is the central object in the construction of Brownian motion, Wiener measure and Feynman path integration.

[24] You should try to prove this fact. It is an easy exercise, just convert to polar coordinates.

Proof. Define

$$u(\tau, z) := \int_{\mathbb{R}^n} p(\tau, z, y) f(y)\, dy, z \in \mathbb{C}^n, \tau = t + is, t > 0, \tag{63}$$

where

$$p(\tau, z, y) = \frac{1}{(4\pi\tau)^{n/2}} e^{-\frac{|z-y|^2}{4\tau}}.$$

To check real analyticity of $u(t, x)$, it suffices to prove complex analyticity of $u(\tau, z)$ on $\mathbb{C}^n \times \operatorname{Re}\tau > 0$. In turn, it suffices to check complex analyticity separately in each variable. Pick a variable z_j and let Γ be a closed contour in the z_j plane. Then,

$$\begin{aligned}\int_\Gamma u(\tau, z) &= \int_{\mathbb{R}^n} \left(\int_\Gamma p(\tau, z, y) f(y)\, dz_j \right) dy \\ &= 0,\end{aligned}$$

since the integrand $p(\tau, z, y)$ is holomorphic in the z_j variable. Now we apply Morera's theorem to see that $u(\tau, z)$ is holomorphic in the z_j variable. The case for τ can be tackled similarly[25]. □

Remark 6.30. *The above phenomenon is quite interesting. Observe that regardless of the support of the initial condition f, the support of $u(t, x)$ is all of $\mathbb{R}^n$, because u is real analytic. This is known in physics as the "infinite speed of propagation of diffusion/Brownian motion". You might have seen something similar in kinetic theory of gases.*

These ideas can be used to approximate L^p functions with smooth functions. With a slight more effort (and using the second part of the above exercise), one can get a quick proof of the Weierstrass approximation theorem. The idea is the same, take an $f \in C([0, 1])$, extend it to a $C_c(\mathbb{R})$ function (which we still call f by minor abuse of notation), and approximate it with a real analytic function (via integrating against the one-dimensional heat kernel). This real analytic function will have a Taylor series expansion, and throw out a tail of the series to get a polynomial to approximate with desired accuracy.

Observe that we have already justified that the Fourier transform is bounded from $L^1 \mapsto L^\infty$ and $L^2 \mapsto L^2$ (the latter is the Plancherel theorem). By the Riesz-Thorin interpolation theorem[26], this implies that the Fourier transform is bounded $L^p \mapsto L^q$, $p \in [1, 2]$, $\frac{1}{p} + \frac{1}{q} = 1$. Observe that since L^p functions are tempered distributions, one can make sense of Fourier transform of L^p functions as tempered distributions[27]. It can however be shown that $\mathcal{F}(L^p(\mathbb{R}^n))$ does not lie inside $L^q(\mathbb{R}^n)$ for $p > 2$[28].

We finish this section with a characterisation of Hilbert-Schmidt operators. The following characterisation of kernels of Hilbert-Schmidt integral operators is used to prove the celebrated Schwartz kernel theorem.

[25] Though at this moment we are not that interested in time real analyticity.

[26] Here is an elementary version of the Riesz-Thorin:

Theorem 6.31 (Riesz-Thorin interpolation)**.** *Let $(\Omega_1, \Sigma_1, \mu_1)$ and $(\Omega_2, \Sigma_2, \mu_2)$ be σ-finite measure spaces. Suppose $1 \leq p_0, q_0, p_1, q_1 \leq \infty$, and let T be a linear operator that boundedly maps $L_{p_0}(\mu_1)$ into $L_{q_0}(\mu_2)$ and $L_{p_1}(\mu_1)$ into $L_{q_1}(\mu_2)$. For $0 < \theta < 1$, let p_θ be defined as*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1},$$

and similarly for q_θ. Then T boundedly maps $L_{p_\theta}(\mu_1)$ into $L_{q_\theta}(\mu_2)$ and satisfies the operator norm estimate

$$\|T\|_{L^{p_\theta} \to L^{q_\theta}} \leq \|T\|^{1-\theta}_{L^{p_0} \to L^{q_0}} \|T\|^{\theta}_{L^{p_1} \to L^{q_1}} . \|T\|_{L^{p_\theta} \to L^{q_\theta}} \leq \|T\|^{1-\theta}_{L^{p_0} \to L^{q_0}} \|T\|^{\theta}_{L^{p_1} \to L^{q_1}} .$$

[27] Have a look at my functional analysis notes if you are interested!

[28] This is not so trivial. One reference is [Ḧ0, Chapter 7, Section 7.6].

Theorem 6.32. *If $T : L^2(X_1, \mu_1) \to L^2(X_2, \mu_2)$ is Hilbert-Schmidt, then there exists a kernel $K \in L^2(X_1 \times X_2, \mu_1 \times \mu_2)$ such that*

$$(Tu, v) = \int \int K(x, y)u(x)\overline{v(y)}d\mu_1(x)d\mu_2(y).$$

Furthermore, we have that

$$\|T\| = \|k\|_{L^2(\mu_1 \times \mu_2)}.$$

The proof is clear by picking ONB u_j for $L^2(X_1, \mu_1)$ and v_k for $L^2(X_2, \mu_2)$, and setting

$$K(x, y) = \sum_{j,k} a_{jk}\overline{u_j(x)}v_k(y),$$

where $a_{jk} = (Tf_j, v_k)$. The rest is left to the reader.

7. Appendix: Bare bones of distribution theory

7.1. Fourier transform and tempered distributions.

Definition 7.1. *A tempered distribution is defined as a continuous linear functional $w : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$. This space is denoted by $\mathcal{S}^*(\mathbb{R}^n)$, and normally considered with the usual weak* topology.*

As previously discussed, for $w : \mathcal{S}(\mathbb{R}^n) \to \mathbb{C}$ to be continuous, it is necessary and sufficient that for some k, C

$$|w(u)| \leq Cp_k(u), \text{ for all } u \in \mathcal{S}(\mathbb{R}^n)$$

(recall that a typical basic open set in $\mathcal{S}(\mathbb{R}^n)$ is a finite intersection of the type $\cap_{j=1}^k p_{i_j}^{-1}(B(0, \epsilon_{i_j}))$. However, we have that if $l \leq m$, then $p_l \leq p_m$, which means that $\cap_{j=1}^k p_{i_j}^{-1}(B(0, \epsilon_{i_j})) = p_{i_k}(B(0, \epsilon_{i_k}))$.

Now we try to indicate some obvious members of the space of tempered distributions. For starters, $\mathcal{S}(\mathbb{R}^n)$ is sitting inside $\mathcal{S}^*(\mathbb{R}^n)$ via the integral pairing

$$f(u) = \int u(x)f(x)dx, \quad u \in \mathcal{S}(\mathbb{R}^n), f \in \mathcal{S}(\mathbb{R}^n). \tag{64}$$

Similarly, $L^p(\mathbb{R}^n) \hookrightarrow \mathcal{S}^*(\mathbb{R}^n)$ for any $p \in [1, \infty]$, via (64), but this time with $f \in L^p(\mathbb{R}^n), u \in \mathcal{S}(\mathbb{R}^n)$.

The basic non-trivial example of a tempered distribution (which is not a function) is the Dirac "delta function" δ[29], defined by

$$\delta(u) = u(0).$$

Now we extend the concept of partial derivatives to the setting of tempered distributions. Each differential operator $D_j = -i\frac{\partial}{\partial x_j}$ acts on $\mathcal{S}^*(\mathbb{R}^n)$ via the definition (which is clearly inspired by the integration by parts formula)

$$D_j w(u) = w(-D_j u), \quad w \in \mathcal{S}^*(\mathbb{R}^n), u \in \mathcal{S}(\mathbb{R}^n).$$

By induction, one can arrive at the formula

$$D^\alpha w(u) = (-1)^{|\alpha|}(w, D^\alpha u).$$

Exercise 7.2. *Prove that $D^\alpha = D^{\alpha_1} \dots D_n^{\alpha_n}$ is a continuous linear operator on $\mathcal{S}^*(\mathbb{R}^n)$.*

[29]If you have studied the concept of "impulse" in physics, this should make sense to you. The Dirac delta models, for example, the situation of a perfectly elastic bat striking an elastic ball. The force acts only for an instant, and it is basically infinite, as the change in velocity is instantaneous.

Exercise 7.3. *Define the Heaviside function on* $\mathbb{R}$ *by*

$$H(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Show that H *lies in* $\mathcal{S}^*(\mathbb{R}^n)$ *and* $\frac{d}{dx}H = \delta$.

Another non-trivial (but very important) example of a tempered distribution is that of a principal valued distribution:

$$\text{p.v.}(u) := \int_0^{+\infty} \frac{u(x) - u(-x)}{x}\, dx. \tag{65}$$

Exercise 7.4. *Prove that (65) above indeed defines a tempered distribution.*

Exercise 7.5. *Find the first distributional derivative of* $\log|x|$.

The beauty of tempered distributions is the natural interaction with Fourier transform, which stems from the corresponding properties of Schwartz spaces. Namely, $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ extends via the formula

$$\mathcal{F}w(u) = w(\mathcal{F}u), \quad w \in \mathcal{S}^*(\mathbb{R}^n), u \in \mathcal{S}(\mathbb{R}^n).$$

Exercise 7.6. *It is straightforward to check that* $\mathcal{F}$ *is a homeomorphism on* $\mathcal{S}^*(\mathbb{R}^n)$. *Do it!*

Observe that we have already justified that the Fourier transform is bounded from $L^1 \mapsto L^\infty$ and $L^2 \mapsto L^2$ (the latter is the Plancherel theorem). By the Riesz-Thorin interpolation theorem, this implies that the Fourier transform is bounded $L^p \mapsto L^q$, $p \in [1, 2]$, $\frac{1}{p} + \frac{1}{q} = 1$. Observe that since L^p functions are tempered distributions, one can make sense of Fourier transform of L^p functions as tempered distributions (via the above definition).

Exercise 7.7. *Prove that* $\mathcal{F}\delta = (2\pi)^{-n/2}$. *(Harder) Prove that*

$$\hat{H}(\xi) = (2\pi)^{-1/2} \lim_{\varepsilon \searrow 0} (\varepsilon + i\xi)^{-1} \text{ in } \mathcal{S}^*(\mathbb{R}^n).$$

Hint: It is not possible to directly calculate the Fourier transform of the Heaviside function. But one can take advantage of the continuity of Fourier transform on $\mathcal{S}^*(\mathbb{R}^n)$ *to define*

$$H_\varepsilon(x) = \begin{cases} e^{-\varepsilon x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0, \end{cases}$$

check that $H_\varepsilon \to H$ *in* $\mathcal{S}(\mathbb{R}^n)$ *as* $\varepsilon \searrow 0$, *which gives the given expression.*

We quickly define the notion of support of a distribution. You will see that it generalises the corresponding notion for functions.

Definition 7.8. *A distribution* w *is said to vanish on an open set* $\Omega \subset \mathbb{R}^n$ *if* $w(u) = 0$ *for all* $u \in C_c^\infty(\Omega)$. *Then, the support of the distribution is the complement of the largest vanishing set of* w.

Of particular interest is the space of compactly supported distributions on $\mathbb{R}^n$. This space is denoted by

$$\mathcal{E}^*(\mathbb{R}^n).$$

Exercise 7.9. *Prove that* $\mathcal{E}^*(\mathbb{R}^n)$ *is precisely the dual space of* $C^\infty(\mathbb{R}^n)$. *(Hint: Observe the fact that any compactly supported distribution can be extended to a continuous linear functional on* C^∞ *by multiplying by a smooth compact cut-off.)*

Also, we state an useful approximation lemma here:

Exercise 7.10. *Prove that $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{S}(\mathbb{R}^n)$. (Hint: Given $u \in \mathcal{S}(\mathbb{R}^n)$, define $\phi_n(x) = \phi(x/n)$, where ϕ is a bump function such that $|\phi| = 1$ on $B_1(0)$, and $\phi = 0$ outside $B_2(0)$. Now try to show that $\phi_n u \to u$ in the Schwartz topology.)*

Now, we come to our first result in the area, which is a (qualitative) variant of the Paley-Weiner theorem.

Proposition 7.11. *Let $w \in \mathcal{E}^*(\mathbb{R}^n)$ be a compactly supported distribution. Then $\mathcal{F}w$ is represented by a smooth function on $\mathbb{R}^n$, given by*

$$\mathcal{F}w(\xi) = \frac{1}{(2\pi)^{n/2}} w(e^{-ix\cdot\xi}).$$

The expression on the right extends to an entire function of $\xi \in \mathbb{C}^n$.

Proof. Since w has compact support, it defines a tempered distribution. Thus, for $u \in \mathcal{S}(\mathbb{R}^n)$,

$$(\mathcal{F}w, u) = (w, \mathcal{F}u).$$

We first prove

$$\left(w, \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} u(\xi) e^{-ix\cdot\xi}\, d\xi\right) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} u(\xi) w(e_\xi)\, d\xi \tag{66}$$

for $u \in C_c^\infty(\mathbb{R}^n)$.

For clarity, consider $n = 1$; the same argument uses multidimensional Riemann sums when $n > 1$. Choose $[a, b]$ containing $\operatorname{supp} u$, set $h = (b-a)/k$, and define

$$v_k(x) = h \sum_{r=1}^{k-1} u(a + rh) e^{-ix(a+rh)}.$$

The Riemann sums v_k converge to

$$v(x) = \int_a^b u(\xi) e^{-ix\xi}\, d\xi$$

uniformly on compact sets, together with every derivative in x. Hence $w(v_k) \to w(v)$. On the other hand,

$$w(v_k) = h \sum_{r=1}^{k-1} u(a + rh) w(e_{a+rh}) \longrightarrow \int_a^b u(\xi) w(e_\xi)\, d\xi.$$

This proves (66) for $u \in C_c^\infty(\mathbb{R})$.

To pass to arbitrary $u \in \mathcal{S}(\mathbb{R}^n)$, choose $\chi \in C_c^\infty(\mathbb{R}^n)$ with $\chi = 1$ near the origin and set $u_R(\xi) = \chi(\xi/R) u(\xi)$. Then $u_R \to u$ in the Schwartz topology, so $\mathcal{F}u_R \to \mathcal{F}u$ in the Schwartz topology and

$$(w, \mathcal{F}u_R) \longrightarrow (w, \mathcal{F}u).$$

Also, compact support of w gives constants C, m such that

$$|w(e_\xi)| \le C(1 + |\xi|)^m.$$

Indeed, choose a compactly supported smooth cutoff equal to one near $\operatorname{supp} w$ and apply the finite-order estimate for w to the cutoff times e_ξ. Since u is rapidly decreasing, dominated convergence yields

$$\int_{\mathbb{R}^n} u_R(\xi) w(e_\xi)\, d\xi \longrightarrow \int_{\mathbb{R}^n} u(\xi) w(e_\xi)\, d\xi.$$

Thus (66) holds for every $u \in \mathcal{S}(\mathbb{R}^n)$ and identifies $\mathcal{F}w$ with the stated function.

Finally, for $z \in \mathbb{C}^n$, define

$$F(z) = \frac{1}{(2\pi)^{n/2}} w(e^{-ix\cdot z}).$$

On compact subsets of $\mathbb{C}^n$, the functions $e^{-ix\cdot z}$ and all their derivatives in x and z are uniformly bounded for x in a fixed neighborhood of $\operatorname{supp} w$. We may therefore differentiate through w, obtaining

$$\partial_{z_j} F(z) = \frac{1}{(2\pi)^{n/2}} w(-ix_j e^{-ix\cdot z}).$$

Hence F is entire and its restriction to $\mathbb{R}^n$ is $\mathcal{F}w$. □

We remark that the definition of $\mathcal{F}^*$ also extends to $\mathcal{S}^*(\mathbb{R}^n)$ in the obvious way:

$$\mathcal{F}^* w(u) = w(\mathcal{F}^*(u)).$$

The following is the dual version of Theorem 6.26:

Proposition 7.12. *We have*

$$\mathcal{F}^*\mathcal{F} = \mathcal{F}\mathcal{F}^* = I \quad \text{on } \mathcal{S}^*(\mathbb{R}^n).$$

Now, we show a rather strong version of Littlewood's second principle:

Theorem 7.13. *$C_c^\infty(\mathbb{R}^n)$ is sequentially dense in $\mathcal{S}^*(\mathbb{R}^n)$in the weak* topology.*

Proof. Pick a $w \in \mathcal{S}^*(\mathbb{R}^n)$ and observe that $\phi_j w \to w$ in the weak*-topology (clear from definition), where ϕ_j is as in Exercise 7.10.

Now, we define the following:

$$w_{jk} := \phi_k(\mathcal{F}^*\phi_j\mathcal{F}w).$$

The first claim is that $w_{jk} \in C_c^\infty(\mathbb{R}^n)$. It suffices to show that

$$\mathcal{F}^*\phi_j\mathcal{F}w \in C^\infty(\mathbb{R}^n). \tag{67}$$

But, $\phi_j\mathcal{F}w \in C_c^\infty(\mathbb{R}^n)$, which implies (67) via application of Proposition 7.11 above.
The next claim is that since $\phi_j\mathcal{F}w \to \mathcal{F}w$, by continuity, we have that $\mathcal{F}^*(\phi_j\mathcal{F}w) \to \mathcal{F}^*\mathcal{F}w$, which implies that $\phi_k(\mathcal{F}^*(\phi_j\mathcal{F}w)) \to \mathcal{F}^*\phi_j\mathcal{F}w \to w$. □

Now we state the following characterization of point-supported tempered distributions.

Theorem 7.14. *If $w \in \mathcal{S}^*(\mathbb{R}^n)$ is supported at $\{0\}$, then there exists k and $a_\alpha \in \mathbb{C}$ such that*

$$w = \sum_{|\alpha|\le k} a_\alpha D^\alpha \delta.$$

Exercise 7.15. *Read the definition of the order of a distribution. Give three examples of distributions of finite order. Can you find an example of a distribution of infinite order?*

A more general object than a tempered distribution is a *distribution*, which is defined as a continuous linear map

$$w : C_c^\infty(\mathbb{R}^n) \to \mathbb{C}.$$

We recall that $C_c^\infty(\mathbb{R}^n)$ is given the so-called "inductive-limit" topology, which is the finest topology on $C_c^\infty(\mathbb{R}^n)$ such that each inclusion $C^\infty(B_{r_k}(0)) \hookrightarrow C_c^\infty(\mathbb{R}^n)$ is continuous for a sequence $r_k \nearrow \infty$[30]. Use the Baire category theorem to convince yourself that with this topology $C_c^\infty(\mathbb{R}^n)$ is not metrizable. It is rather difficult to work directly with this space, but duality makes things easier, which was the idea of distributions and one of the recurring themes in the work of Hörmander. It can be proved[31] that continuity is equivalent to the following:

[30] It is straightforward to check that this does not depend on the specific sequence r_k.

[31] If you are interested, you can ask me for a reference, which is going to be one of the volumes of Hörmander!

Theorem 7.16. *A linear map $C_c^\infty(\mathbb{R}^n)$ is continuous iff for any sequence $\varphi_n \in C_c^\infty(\mathbb{R}^n)$ which is supported inside a common compact $K \subset \mathbb{R}^n$ and satisfies that there exists $\varphi \in C_c^\infty(\mathbb{R}^n)$ such that φ_n converges to φ in C^∞, then*

$$w(\varphi_n) \to w(\varphi).$$

It is of interest to define the Fourier transform of a distribution. Given $w \in \mathcal{D}^*(\mathbb{T}^n)$, define the following Fourier series:

$$\mathcal{F}w(k) = \frac{1}{(2\pi)^{n/2}} w(e^{-ik.\theta}).$$

Since w must satisfy as estimate of the form

$$|w(u)| \leq C\|u\|_{C^l(\mathbb{T}^n)}.$$

it is clear that

$$\mathcal{F} : \mathcal{D}^*(\mathbb{T}^n) \to s^*(\mathbb{Z}^n),$$

where

$$s^*(\mathbb{Z}^n) = \{a : \mathbb{Z}^n \to \mathbb{C} : |a(k)| \leq C\langle k\rangle^N \text{ for some } C, N\}$$

consists of polynomially bounded functions on $\mathbb{Z}^n$.

It can be checked that $s^*(\mathbb{Z}^n)$ is the dual space to

$$s(\mathbb{Z}^n) := \{u : \mathbb{Z}^n \to \mathbb{C} : p_N(u) = \sup_{k\in\mathbb{Z}^n} \langle k\rangle^N |u(k)| < \infty\}.$$

Also, $\mathcal{F} : \mathcal{D}^*(\mathbb{T}^n) \to s^*(\mathbb{Z}^n)$ is the adjoint of the map $\mathcal{F}^* : s(\mathbb{Z}^n) \to C^\infty(\mathbb{T}^n)$ defined by

$$\mathcal{F}^*u(\theta) = \sum_{k\in\mathbb{Z}^n} u(k)e^{ik.\theta}.$$

The Fourier inversion formulas

$$\mathcal{F}^*\mathcal{F} = I \text{ on } C^\infty(\mathbb{T}^n), \quad \mathcal{F}\mathcal{F}^* = I \text{ on } s(\mathbb{Z}^n),$$

extend by duality to

$$\mathcal{F}^*\mathcal{F} = I \text{ on } \mathcal{D}^*(\mathbb{T}^n), \quad \mathcal{F}\mathcal{F}^* = I \text{ on } s^*(\mathbb{Z}^n).$$

8. Radon measures and the Riesz-Kakutani represention theorem

For this section, we shall follow almost verbatim the treatment in [Fol99]. Let X be a locally compact Hausdorff space, and let $\mathcal{B}$ denote the Borel σ-algebra. Also, recall that a linear functional T on $C_c(X)$ is called positive if $T(f) \geq 0$ whenever $f \geq 0$. Also, throughout this section, $\|f\|$ means the supremum of f unless mentioned otherwise. We leave the following as an exercise:

Exercise 8.1. *If T is a positive linear functional on $C_c(X)$, for each compact $K \subseteq X$ there is a constant C_K such that $|T(f)| \leq C_K\|f\|$ for all $f \in C_c(X)$ such that $\operatorname{supp}(f) \subseteq K$. (Hint: Use Urysohn's lemma to find a $C_c(X)$ function φ taking values in $[0,1]$ that is identically 1 on K. Then, consider the non-negative functions $\|f\|\varphi \pm f$).*

If μ is a Borel measure on X such that $\mu(K) < \infty$ for every compact $K \subseteq X$, then clearly $C_c(X) \subseteq L^1(X,\mu)$, so the map $f \mapsto \int f\, d\mu$ is a positive linear functional on $C_c(X)$. The Riesz-Kakutani representation theorem says that *every* positive linear functional on $C_c(X)$ arises this way. We introduce the following definitions. Recall that a Borel measure is called *outer regular* on a Borel set E if $\mu(E) = \inf\{\mu(U) : E \subseteq U, U \text{ open }\}$, and *inner regular* on E if $\mu(E) = \sup\{\mu(K) : K \subseteq E, K \text{ compact }\}$. If the above happen for all Borel sets, μ is called *regular*. A *Radon measure* on X is a Borel measure that is finite on all compact sets, outer regular on all Borel sets and inner regular on all open sets.

Further, as a notation, we use $f \prec U$ to mean that $0 \leq f \leq 1, \operatorname{supp} f \subseteq U$, where $f \in C_c(X)$ and $U \subseteq X$ is open.

Theorem 8.2 (Riesz-Kakutani for positive functionals)**.** *If T is a positive linear functional on $C_c(X)$, then there is a unique Radon measure μ on X such that $T(f) = \int f\, d\mu$ for all $f \in C_c(X)$. Moreover, μ satisfies*

$$\mu(U) = \sup\{T(f) : f \prec U\} \text{ for all open } U, \tag{68}$$

and

$$\mu(K) = \inf\{T(f) : f \in C_c(X), f \geq \chi_K\} \text{ for all compact } K \subseteq X. \tag{69}$$

Consequently, every positive bounded linear functional T on $C_0(X)$ is given by integration against finite Radon measures.

Proof. We first discuss the issue of uniqueness. Suppose we are able to find a Radon measure μ satisfying all the conditions above. Then, for any $f \prec U$, one sees that $T(f) \leq \mu(U)$. On the other hand, by applying Urysohn's lemma to the pair $K, X \setminus U$, where K is an arbitrary compact set inside U, one can find a $f \in C_c(X)$ such that $T(f) \geq \mu(K)$. Because of inner regularity on U, one sees that μ is uniquely determined on all open sets. Because of outer regularity, this fixes μ.

Now, the issue of existence. Begin by defining

$$\mu(U) = \sup\{T(f) : f \prec U\} \text{ for open } U, \tag{70}$$

and then define $\mu^*(E)$ for an arbitrary $E \subseteq X$ by

$$\mu^*(E) = \inf\{\mu(U) : E \subseteq U, U \text{ open }\}.$$

Clearly, $\mu(U) \leq \mu(V)$ if $U \subseteq V$ and $\mu^*(U) = \mu(U)$ if U is open.
Step I: We establish that μ^* is an outer measure. It suffices to show that if U_j is a sequence of open sets and $U = \bigcup_j U_j$, then $\mu(U) \leq \sum_j \mu(U_j)$. It follows from this that for any $E \subseteq X$,

$$\mu^*(E) = \inf\left\{\sum_j \mu(U_j) : U_j \text{ open}, E \subseteq \bigcup_j U_j\right\},$$

and the expression on the right defines an outer measure (this follows from the construction in Section 2, in particular, Theorem 2.2). Let $f \prec U$ with compact support K. Since K is compact, we have that $K \subseteq \bigcup_{j=1}^n U_j$ for some finite n, so there exists a partition of unity g_j subordinate to this finite open cover. But then, $f = \sum_{j=1}^n f g_j$ and $f g_j \prec U_j$, so

$$T(f) = \sum_{j=1}^n T(f g_j) \leq \sum_{j=1}^n \mu(U_j) \leq \sum_{j=1}^\infty \mu(U_j).$$

Since this is true for any $f \prec U$, we conclude that $\mu(U) \leq \sum_{j=1}^\infty \mu(U_j)$, as desired.
Step II: We now establish that every open set is μ^*-measurable. It suffices to show that if $\mu^*(E) < \infty$ then $\mu^*(E) \geq \mu^*(E \cap U) + \mu^*(E \setminus U)$. First suppose that E is open. Then $E \cap U$ is open, so given $\varepsilon > 0$ we can find $f \in C_c(X)$ such that $f \prec E \cap U$ and $T(f) > \mu(E \cap U) - \varepsilon$. Also, $E \setminus \operatorname{supp} f$ is open, so we can find $g \prec E \setminus \operatorname{supp} f$ and $T(g) > \mu(E \setminus \operatorname{supp} f) - \varepsilon$. Then, $f + g \prec E$, so

$$\begin{aligned}\mu^*(E) &\geq T(f) + T(g) > \mu(E \cap U) + \mu(E \setminus \operatorname{supp} f) - 2\varepsilon \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U) - 2\varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, we obtain the desired inequality. For the general case, if $\mu^*(E) < \infty$ we can find an open $V \supseteq E$ such that $\mu(V) < \mu^*(E) + \varepsilon$, and hence

$$\begin{aligned}\mu^*(E) + \varepsilon > \mu(V) &\geq \mu^*(V \cap U) + \mu^*(V \setminus U) \\ &\geq \mu^*(E \cap U) + \mu^*(E \setminus U).\end{aligned}$$

Letting $\varepsilon \to 0$, we are done.

Step III: We now show that μ satisfies (69). If K is compact, $f \in C_c(X), f \geq \chi_K$, let $U_\varepsilon = \{x : f(x) > 1 - \varepsilon\}$. Then U_ε is open, and if $g \prec U_\varepsilon$, we have $(1-\varepsilon)^{-1}f - g \geq 0$, which implies that $T(g) \leq (1-\varepsilon)^{-1}T(f)$. Thus $\mu(K) \leq \mu(U_\varepsilon) \leq (1-\varepsilon)^{-1}T(f)$. Letting $\varepsilon \to 0$, we see that $\mu(K) \leq T(f)$. Also, for any open $U \supseteq K$, by Urysohn's lemma there exists $f \in C_c(X)$ such that $f \geq \chi_K$ and $f \prec U$, whence $T(f) \leq \mu(U)$. Since μ is outer regular on K, (69) is proved.

Step IV: Now, we prove that $T(f) = \int f \, d\mu$ for all $f \in C_c(X)$. It suffices to show that $T(f) = \int f \, d\mu$ if $f \in C_c(X, [0,1])$, as $C_c(X)$ is the linear span of the latter set. Let $K_j = \{x : f(x) \geq j/N\}, 1 \leq j \leq N$ and let $K_0 = \operatorname{supp} f$. Also, define $f_1, \ldots, f_N \in C_c(X)$ by $f_j(x) = 0$ if $x \notin K_{j-1}, f_j(x) = f(x) - (j-1)/N$ if $x \in K_{j-1} \setminus K_j$, and $f_j(x) = 1/N$ if $x \in K_j$. In other words,

$$f_j = \min\left\{\max\left\{f - \frac{j-1}{N}, 0\right\}, \frac{1}{N}\right\}.$$

Then, $\chi_{K_j}/N \leq f_j \leq \chi_{K_{j-1}}/N$, hence

$$\frac{1}{N}\mu(K_j) \leq \int f_j \, d\mu \leq \frac{1}{N}\mu(K_{j-1}).$$

Also, if U is an open set containing K_{j-1} we have $Nf_j \prec U$ and so $T(f_j) \leq \mu(U)/N$. Hence by (69) and outer regularity,

$$\frac{1}{N}\mu(K_j) \leq T(f_j) \leq \frac{1}{N}\mu(K_{j-1}).$$

Moreover, $f = \sum_1^N f_j$, so that

$$\frac{1}{N}\sum_1^N \mu(K_j) \leq \int f \, d\mu \leq \frac{1}{N}\sum_0^{N-1} \mu(K_j),$$

$$\frac{1}{N}\sum_1^N \mu(K_j) \leq T(f) \leq \frac{1}{N}\sum_0^{N-1} \mu(K_j).$$

It follows that

$$|T(f) - \int f \, d\mu| \leq \frac{\mu(K_0) - \mu(K_N)}{N} \leq \frac{\mu(\operatorname{supp} f)}{N}.$$

Since $\mu(\operatorname{supp} f) < \infty$ and N is arbitrary, we conclude that $T(f) = \int f \, d\mu$.

Step V: Since $C_0(X)$ is the uniform closure of $C_c(X)$[32], the functional $T(f) = \int f \, d\mu$ extends continuously to $C_0(X)$ iff it is bounded with respect to the uniform norm. In view of the equality (special case of (70))

$$\mu(X) = \sup\left\{\int_X f \, d\mu : f \in C_c(X), 0 \leq f \leq 1\right\}$$

together with the fact that $|\int f \, d\mu| \leq \int |f| \, d\mu$, this happens precisely when $\mu(X) < \infty$, in which case $\mu(X)$ is the operator norm of T. □

Now, we extend the above result to identify the bounded linear functionals (in other words, the dual space) of $C_0(X)$. We start with the following

Lemma 8.3. *If $T \in C_0(X, \mathbb{R})^*$, there exist positive functionals $T^\pm \in C_0(X, \mathbb{R})^*$ such that $T = T^+ + T^-$.*

Since this has very little to do with measure theory, we shall only sketch the basic idea of the proof, and leave it to the reader to check the details.

[32]Let $f \in C_0(X)$ and $\varepsilon > 0$. The set $K = \{x \in X : |f(x)| \geq \varepsilon\}$ is compact. By Urysohn's lemma, choose $\chi \in C_c(X)$ with $0 \leq \chi \leq 1$ and $\chi = 1$ on K. Then $\chi f \in C_c(X)$ and $\|f - \chi f\|_\infty \leq \varepsilon$.

Sketch. If $f \in C_0(X, [0,\infty))$, define $T^+(f) = \sup\{T(g) : g \in C_0(X,\mathbb{R}), 0 \le g \le f\}$. So, if $f \in C_0(X,\mathbb{R})$, then its positive and negative parts f^+ and f^- are in $C_0(X,[0,\infty))$, and we define $T^+(f) = T^+(f^+) - T^+(f^-)$. Finally, we define $T^- = T^+ - T$.

□

Recall that we talked about complex Radon measures already. Let the space of complex Radon measures on X be denoted by $\mathfrak{M}(X)$. For $\mu \in \mathfrak{M}(X)$, we define

$$\|\mu\| = |\mu|(X), \tag{71}$$

where $|\mu|$ is the total variation of μ. Recall that for a complex measure μ, its *variation* is defined as the following measure:

$$|\mu|(A) = \sup \sum_{j\ge 1} |\mu(A_j)|, \tag{72}$$

where the supremum ranges over all sequences of disjoint sets A_j whose union is A. We leave the following as an exercise:

Exercise 8.4. *If μ is a complex Borel measure, then μ is Radon iff $|\mu|$ is Radon. Moreover, $\mathfrak{M}(X)$ is a vector space and $\mu \mapsto \|\mu\|$ is a norm.*

Now we come to our main goal:

Theorem 8.5 (Riesz-Kakutani representation)**.** *Let X be a locally compact Hausdorff space and let $T_\mu(f) = \int f \, d\mu, f \in C_0(X), \mu \in \mathfrak{M}(X)$. Then the map $\mu \mapsto T_\mu$ is an isometric isomorphism from $\mathfrak{M}(X)$ to $C_0(X)^*$.*

Proof. We have already shown that every $T \in C_0(X)^*$ is of the form T_μ. On the other hand, if $\mu \in \mathfrak{M}(X)$, we have that

$$\left|\int f \, d\mu\right| \le \int |f| \, d|\mu| \le \|f\|_{L^\infty}\|\mu\|,$$

so $T_\mu \in C_0(X)^*$ and $\|T_\mu\| \le \|\mu\|$. Moreover if $h = d\mu/d|\mu|$, then $|h| = 1$, so by Lusin's theorem, for any $\varepsilon > 0$, there exists $f \in C_c(X)$ such that $\|f\|_{L^\infty} = 1$ and $f = \overline{h}$ except on a set E with $|\mu|(E) < \frac{\varepsilon}{2}$. Then

$$\begin{aligned}\|\mu\| = \int |h|^2 \, d|\mu| = \int \overline{h} \, d\mu &\le \left|\int f \, d\mu\right| + \left|\int (f - \overline{h}) \, d\mu\right| \\ &\le \left|\int f \, d\mu\right| + 2|\mu|(E) < \left|\int f \, d\mu\right| + \varepsilon \le \|T_\mu\| + \varepsilon.\end{aligned}$$

It follows that $\|\mu\| \le \|T_\mu\|$, so the proof is complete. □

8.1. **A few remarks and immediate consequences.** For a more "topological" proof of the Riesz-Kakutani representation theorem via the Stone-Čech compactification, have a look at [Gar73]. There is a "categorical formulation" in [Har83] which also includes a description of the former paper.

We shall have time to apply the Riesz-Kakutani theorem in detail later, but here we note down a few immediate consequences of Theorem 8.5. Then, the Banach-Alaoglou theorem implies that the closed unit ball in $\mathfrak{M}(X)$ is weak*-compact. An important subset of $\mathfrak{M}(X)$ is the set of probability measures. Recall that an element μ of $\mathfrak{M}(X)$ is called a probability measure if μ is positive and $\mu(X) = 1$. We use the notation $\mathrm{Prob}(X) = \{\mu \in \mathfrak{M}(X) : \mu \ge 0, \mu(X) = 1\}$. So, the upshot is that the collection of probability measures on X is compact in the weak* topology.

9. Sobolev spaces and almost everywhere differentiability

We first introduce the concept of a *weak derivative*. Let $u \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then, its distributional derivative $\partial_j u$ is called a weak derivative if $\partial_j u \in L^1_{\text{loc}}(\mathbb{R}^n)$.

We define the Sobolev space

$$H^{1,p}(\mathbb{R}^n) = \{u \in L^p(\mathbb{R}^n) : \partial_j u \in L^p(\mathbb{R}^n), j = 1, \dots, n\}. \tag{73}$$

Remark 9.1. *We did not separately state that $\partial_j u \in L^1_{\text{loc}}$: on $\mathbb{R}^n$, every L^p function with $1 \le p \le \infty$ is locally integrable. Weak derivatives need not agree with what one obtains by differentiating almost everywhere. For example, $|x|$ has weak derivative $\operatorname{sgn} x$. The function $\operatorname{sgn} x$ is classically differentiable almost everywhere, with derivative zero there, but its distributional derivative is $2\delta_0$. Thus $\operatorname{sgn} x$ has no weak derivative belonging to L^1_{loc}.*

Exercise 9.2. *Let $1 \le p < \infty$. Prove that $H^{1,p}(\mathbb{R}^n)$ is a Banach space with norm*

$$\|u\|_{H^{1,p}} = \|u\|_{L^p} + \sum_{j=1}^{n} \|\partial_j u\|_{L^p}. \tag{74}$$

Hint: *Let (u_k) be Cauchy in this norm. Then (u_k) and each $(\partial_j u_k)$ are Cauchy in L^p. By completeness of L^p, there are functions $u, v_1, \dots, v_n \in L^p$ such that*

$$u_k \longrightarrow u \quad \text{and} \quad \partial_j u_k \longrightarrow v_j \qquad \text{in } L^p.$$

It remains to show that v_j is the weak derivative of u. Fix $\varphi \in C_c^\infty(\mathbb{R}^n)$ and pass to the limit in

$$\int_{\mathbb{R}^n} u_k \, \partial_j \varphi \, dx = -\int_{\mathbb{R}^n} (\partial_j u_k)\varphi \, dx.$$

Use Hölder's inequality to justify both limits. Conclude that $u \in H^{1,p}(\mathbb{R}^n)$ and $u_k \to u$ in the stated norm.

Exercise 9.3. *Derive the following estimates:*

$$\|\varphi * u\|_{H^{1,p}} \le \|\varphi\|_{L^1} \|u\|_{H^{1,p}}, \tag{75}$$

$$\|\psi u\|_{H^{1,p}} \le \|\psi\|_{L^\infty} \|u\|_{H^{1,p}} + \sum_j \|\partial_j \psi\|_{L^\infty} \|u\|_{L^p}, \tag{76}$$

for $\varphi, \psi \in C_c^\infty(\mathbb{R}^n), u \in H^{1,p}(\mathbb{R}^n)$.

Exercise 9.4. *Prove that $C_c^\infty(\mathbb{R}^n)$ is dense in $H^{1,p}(\mathbb{R}^n)$ for $p \in [1, \infty)$. (Hint: take convolution with peaked bump functions to approximate by smooth functions, and then multiply with a smooth cut-off to get compact support).*

The following is a version of the Sobolev embedding theorem.

Theorem 9.5 (Sobolev embedding)**.** *Let $1 \le p < \infty$. If $p > n$, or if $p = n = 1$, then every $u \in H^{1,p}(\mathbb{R}^n)$ has a bounded continuous representative. In particular,*

$$H^{1,p}(\mathbb{R}^n) \subseteq C(\mathbb{R}^n) \cap L^\infty(\mathbb{R}^n). \tag{77}$$

Remark 9.6. *The inclusion (77) means that every equivalence class in $H^{1,p}(\mathbb{R}^n)$ contains a bounded continuous function. This representative agrees almost everywhere with the original function, or equivalently defines the same distribution.*

Proof. We first prove that

$$\|u\|_{L^\infty} \le C\|u\|_{H^{1,p}}, \qquad u \in C_c^\infty(\mathbb{R}^n).$$

By translation, it suffices to bound $|u(0)|$ by the right-hand side.

Write $B_r = B(0,r)$ and let ω_n be the volume of B_1. For $y \in B_r$, the fundamental theorem of calculus gives

$$u(0) = u(y) - \int_0^1 \sum_{j=1}^n y_j \partial_j u((1-t)y)\, dt.$$

Integrating over B_r, we obtain

$$\omega_n r^n |u(0)| \le \int_{B_r} |u(y)|\, dy + \sum_{j=1}^n \int_0^1 \int_{B_r} |y|\, |\partial_j u((1-t)y)|\, dy\, dt.$$

For the last integral, set $z = (1-t)y$ and then change the order of integration. It becomes

$$\int_{B_r} |z|\, |\partial_j u(z)| \int_0^{1-|z|/r} (1-t)^{-n-1}\, dt\, dz.$$

Since

$$\int_0^{1-|z|/r} (1-t)^{-n-1}\, dt = \frac{1}{n}\left(\frac{r^n}{|z|^n} - 1\right),$$

we conclude that

$$|u(0)| \le \frac{1}{\omega_n r^n} \int_{B_r} |u(y)|\, dy + \frac{1}{n\omega_n} \sum_{j=1}^n \int_{B_r} |\partial_j u(z)|\, |z|^{1-n}\, dz. \tag{78}$$

Take $r = 1$. If $p > n$, Hölder's inequality bounds the right-hand side by $C\|u\|_{H^{1,p}}$, because

$$\int_{B_1} |z|^{(1-n)p'}\, dz < \infty, \qquad \frac{1}{p} + \frac{1}{p'} = 1,$$

is equivalent to $p > n$. If $p = n = 1$, then $|z|^{1-n} = 1$, and (78) gives the same bound directly.

Applying this estimate to translates of u proves the asserted L^∞ estimate. Now let $u \in H^{1,p}(\mathbb{R}^n)$ and choose $u_k \in C_c^\infty(\mathbb{R}^n)$ with $u_k \to u$ in $H^{1,p}$. The estimate applied to $u_k - u_\ell$ shows that (u_k) is uniformly Cauchy. Its uniform limit is bounded and continuous. Since $u_k \to u$ also in L^p, this limit agrees almost everywhere with u. □

Exercise 9.7. *Show that the proof above can be easily modified to prove that $H^{m+j,p}(\mathbb{R}^n)$ embeds continuously in $C^j(\mathbb{R}^n)$ when $mp > n$ or $m = n, p = 1$.*

Let X be a compact metric space, and introduce the Hölder spaces

$$C^{0,\alpha}(X) = \{u \in C(X) : |u(x) - u(y)| \le C d(x,y)^\alpha\}.$$

Define the norm

$$\|u\|_{C^{k,\alpha}} = \|u\|_{C^k} + \sup_{x,y \in X} \frac{|u(x) - u(y)|}{d(x,y)^\alpha}.$$

Exercise 9.8. *Check that this is a Banach space. (Hint: Again, if you feel like, do it for $X = \overline{\Omega}$, where $\Omega \subset \mathbb{R}^n$ is a bounded open set first. Then it is easy to abstract the proof.)*

Remark 9.9. *The Hölder spaces (along with Sobolev spaces) are of significant importance to analysis, as these are the Banach spaces that interact nicely with elliptic operators. It can be easily shown that (contrary to common sense expectation!) the usual C^k spaces are not well-behaved for elliptic regularity. In fact, one can produce examples of functions u such that $\Delta u \in C(\mathbb{R}^n)$, but $u \notin C^2(\mathbb{R}^n)$[33]. In practise, when u solves an elliptic PDE, oftentimes the hard part in regularity is to get $u \in C^\alpha$ for some α. If the PDE is not too wild, Schauder bootstrapping gives $u \in C^{2,\alpha}$ (if the PDE is of second order), and then inductively one can raise the regularity.*

[33]Exercise: Find an example of such a function!

It goes without saying that elliptic regularity is one of the most pervasive phenomena in Mathematics. It shows up practically in any branch of Mathematics that I know of. In particular, it is at the heart of why analytic functions (or harmonic functions) are infinitely differentiable, finding harmonic cohomology classes in Hodge theory, and the theory of Laplace eigenfunctions, which is the study of harmonic analysis on Riemannian manifolds.

Exercise 9.10 (Now we have done this in class!)**.** *The above proof can also be modified to prove that $H^{1,p}(\mathbb{R}^n)$ satisfies a Hölder condition:*

$$H^{1,p}(\mathbb{R}^n) \subseteq C^s(\mathbb{R}^n), s = 1 - \frac{n}{p}.$$

Here is another useful version of the Sobolev embedding:

Proposition 9.11. *We have the identity*

$$H^{1,\infty}(\mathbb{R}^n) = \mathrm{Lip}(\mathbb{R}^n), \tag{79}$$

where Lip denotes the bounded Lipschitz functions.

Proof. Let u be bounded and Lipschitz. For each j, the difference quotients

$$\frac{u(\,\cdot + he_j) - u}{h}$$

are bounded in $L^\infty(\mathbb{R}^n)$ by the Lipschitz constant of u. Choose $h_k \to 0$ such that these quotients converge weak-* to some $f_j \in L^\infty(\mathbb{R}^n)$. For every $\varphi \in C_c^\infty(\mathbb{R}^n)$, it follows that

$$\int_{\mathbb{R}^n} f_j \varphi \, dx = \lim_{k\to\infty} \frac{1}{h_k} \int_{\mathbb{R}^n} \varphi(x)\big(u(x + h_k e_j) - u(x)\big)\, dx.$$

Changing variables in the first term on the right, and then passing to the limit, gives

$$\int_{\mathbb{R}^n} f_j \varphi \, dx = -\int_{\mathbb{R}^n} u\, \partial_j \varphi \, dx.$$

Hence $\partial_j u = f_j \in L^\infty(\mathbb{R}^n)$, and $u \in H^{1,\infty}(\mathbb{R}^n)$.

Conversely, let $u \in H^{1,\infty}(\mathbb{R}^n)$. Choose $\varphi \in C_c^\infty(\mathbb{R}^n)$ with $\varphi \geq 0$ and $\int \varphi = 1$, and set $\varphi_k(x) = k^n \varphi(kx)$ and $u_k = \varphi_k * u$. Since

$$\partial_j u_k = \varphi_k * \partial_j u,$$

the functions u_k are uniformly bounded and have a common Lipschitz constant L.

At every Lebesgue point x of u, we have $u_k(x) \to u(x)$. Suppose $\operatorname{supp}\varphi \subseteq B(0,R)$. At every Lebesgue point x of u,

$$\begin{aligned} |(\varphi_k * u)(x) - u(x)| &\leq \|\varphi_k\|_\infty \int_{B(x,R/k)} |u(y) - u(x)|\, dy \\ &\leq C_\varphi \fint_{B(x,R/k)} |u(y) - u(x)|\, dy \longrightarrow 0, \end{aligned}$$

where $\int \varphi = 1$. Thus $u_k(x) \to u(x)$ on a set of full measure E. Therefore, for $x, y \in E$,

$$|u(x) - u(y)| = \lim_{k\to\infty} |u_k(x) - u_k(y)| \leq L|x - y|.$$

Thus $u|_E$ extends uniquely to a bounded Lipschitz function on $\mathbb{R}^n$, which agrees with u almost everywhere.[34] □

We quickly digress to define the convolution of a function and a distribution.

[34] We could have avoided the Lebesgue point argument by simply using pointwise convergence up to a subsequence.

Definition 9.12. *Let $w \in \mathcal{D}^*(\mathbb{R}^n), u, \varphi \in C_c^\infty(\mathbb{R}^n)$. Then define the convolution $\varphi * w$ as the distribution which acts like*

$$(\varphi * w, u) = (w, R\varphi * u), \text{ where } R\varphi(x) = \varphi(-x).$$

For this definition to sink in, the reader should check that this is true if w is replaced by another $C_c^\infty(\mathbb{R}^n)$ function.

As a parting remark, you can locate an entire spectrum of Sobolev embeddings in [AF03]. See for example, Theorem 4.12 of loc.cit. Of particular importance are the L^2-based Sobolev spaces, of which we state a few here without proof:

Theorem 9.13 (Continuous Sobolev embedding)**.** *If $s > \frac{n}{2} + k$, we have that*

$$H^{s,2}(\mathbb{R}^n) \subset C^k(\mathbb{R}^n), \tag{80}$$

and the embedding is continuous.

Theorem 9.14 (Compact Sobolev/Rellich embedding)**.** *If $\Omega \subset \mathbb{R}^n$ is a bounded domain, then $H_0^{t,2}(\Omega) \subset H_0^{s,2}(\Omega)$ is a compact embedding*[35] *when $t > s$.*

Proof. We will pick a sequence $f_k \in H_0^{t,2}$ such that $\|f\|_{H_0^{t,2}} \leq C < \infty$. We have to show that f_k has a convergent subsequence in $H_0^{s,2}(\Omega)$.

Choose $\psi \in C_c^\infty(\mathbb{R}^n)$ such that $\psi \equiv 1$ on Ω. Then, $f_k = \psi f_k$, which gives that $\hat{f}_k = \hat{\psi} * \hat{f}_k$, which is a smooth function. We use the following inequality, which is easy to verify:

$$\left(\frac{1+|\xi|^2}{1+|\eta|^2}\right)^t \leq 2^{|t|}(1+|\xi-\eta|^2)^{|t|}.$$

Now, we see that

$$\begin{aligned}(1+|\xi|^2)^{t/2}|\hat{f}_k(\xi)| &\leq 2^{|t|/2}\int |\hat{\psi}(\xi-\eta)|(1+|\xi-\eta|^2)^{|t|/2}|\hat{f}_k(\eta)|(1+|\eta|^2)^{t/2}d\eta \\ &\leq 2^{|t|/2}\|\psi\|_{H^{|t|,2}}\|f_k\|_{H^{t,2}}.\end{aligned}$$

Similarly, one gets an estimate on the derivative,

$$(1+|\xi|^2)^{t/2}|D_j\hat{f}_k(\xi)| \leq 2^{|t|/2}\|x_j\psi\|_{H^{|t|,2}}\|f_k\|_{H^{t,2}}.$$

Since $\|f_k\|_{H^{t,2}} \leq C$, the functions $\hat{f}_k$ and their derivatives are uniformly bounded on compact sets, so by the Arzelà-Ascoli theorem there is a subsequence $\hat{f}_k$ (minor abuse of notation) that converges uniformly on compact sets.

Now, we have the following:

$$\begin{aligned}\|f_m - f_l\|_{H_0^{s,2}}^2 &= \int_{|\xi|\leq R}(1+|\xi|^2)^s|\hat{f_m} - \hat{f_l}|^2(\xi)d\xi + \int_{|\xi|\geq R}(1+|\xi|^2)^{s-t}(1+|\xi|^2)^s|\hat{f_m}\ \hat{f_l}|^2(\xi)d\xi \\ &\leq \sup_{|\xi|\leq R}|\hat{f_m} - \hat{f_l}|^2(\xi)\int_{|\xi|\leq R}(1+|\xi|^2)^s d\xi + (1+R^2)^{s-t}\int_{|\xi|>R}(1+|\xi|^2)^t|\hat{f}_m - \hat{f}_l|^2(\xi)d\xi \\ &= \sup_{|\xi|\leq R}|\hat{f_m} - \hat{f_l}|^2(\xi)\int_{|\xi|\leq R}(1+|\xi|^2)^s d\xi + (1+R^2)^{s-t}\|f_l - f_m\|_{H^{t,2}}^2.\end{aligned}$$

Now we can make both terms arbitrarily small by choosing R sufficiently large, and the first term is arbitrarily small for large enough m, l. This finishes the proof. □

[35] Here, one of the equivalent definitions of $H_0^{s,2}(\Omega)$ is the closure of $C_c^\infty(\Omega)$ in $H^{s,2}(\mathbb{R}^n)$.

We have already seen that almost every point in $\mathbb{R}^n$ is a Lebesgue point for $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

For $f \in L^p(\mathbb{R}^n)$, $1 \le p < \infty$, define

$$\widetilde{M}_p f(x)^p = \sup_{r>0} \frac{1}{|B_r|} \int_{B_r(x)} |f(y) - f(x)|^p \, dy.$$

The following is left as an exercise:

Proposition 9.15. *For every $\lambda > 0$,*

$$\big|\{x \in \mathbb{R}^n : \widetilde{M}_p f(x) > \lambda\}\big| \lesssim_{n,p} \lambda^{-p} \|f\|_{L^p}^p.$$

Moreover, for almost every $x \in \mathbb{R}^n$,

$$\lim_{r \downarrow 0} \frac{1}{|B_r|} \int_{B_r(x)} |f(y) - f(x)|^p \, dy = 0. \tag{81}$$

Proof. Let M be the centered Hardy–Littlewood maximal operator. The inequality

$$|f(y) - f(x)|^p \le 2^{p-1}\big(|f(y)|^p + |f(x)|^p\big)$$

gives

$$\widetilde{M}_p f(x)^p \le 2^{p-1}\big(M(|f|^p)(x) + |f(x)|^p\big).$$

The weak $(1,1)$ estimate for M and Chebyshev's inequality prove the asserted maximal estimate.

We give the argument for the second assertion, since it does not follow from the maximal estimate alone without an approximation step. Set

$$A_f(x) = \limsup_{r \downarrow 0} \frac{1}{|B_r|} \int_{B_r(x)} |f(y) - f(x)|^p \, dy.$$

Fix $g \in C_c^\infty(\mathbb{R}^n)$. Let $h = f - g$. Since g is continuous, the triangle inequality gives

$$A_f(x)^p \le 3^{p-1}\big(M(|h|^p)(x) + |h(x)|^p\big).$$

Let x be a Lebesgue point of h. Then

$$\lim_{r \downarrow 0} \fint_{B(x,r)} h(y) \, dy = h(x).$$

For each $r > 0$, the function $z \mapsto |z|^p$ is convex, so Jensen's inequality applied with the measure $\frac{dy}{|B(x,r)|}$ gives

$$\left| \fint_{B(x,r)} h(y) \, dy \right|^p \le \fint_{B(x,r)} |h(y)|^p \, dy \le M(|h|^p)(x).$$

The right-hand side does not depend on r. Letting $r \downarrow 0$ in the left-hand side, and using continuity of $z \mapsto |z|^p$, we obtain

$$|h(x)|^p \le M(|h|^p)(x).$$

Consequently,

$$A_f(x)^p \le 2 \cdot 3^{p-1} M(|f - g|^p)(x)$$

for almost every x. As $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, the right-hand side can be made arbitrarily small. Thus $A_f = 0$ almost everywhere. □

Parallel to what happened above, call the points where (81) holds L^p Lebesgue points of f.

Also, we recall at this point the notion of classical (pointwise) differentiation. Let $\Omega \subseteq \mathbb{R}^n$ be a domain and $f : \Omega \to \mathbb{R}^m$. Then f is differentiable at a point $x \in \Omega$ provided there is a linear transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$|f(x + y) - f(x) - Ay| = o(|y|), \text{ as } |y| \to 0.$$

If such a linear transformation exists, it must be unique and we denote it by $Df(x)$.

Theorem 9.16. *Let $f \in H^{1,p}(\mathbb{R}^n)$, where either $n < p < \infty$ or $p = n = 1$, and take the continuous representative of f given by the Sobolev embedding theorem. If x is an L^p Lebesgue point of ∇f, then f is differentiable at x and*

$$Df(x) = \nabla f(x).$$

In particular, f is differentiable almost everywhere.

Proof. We first derive the local estimate needed in the proof. Recall that (78), translated to a ball $B(a,\rho)$, states that

$$|u(a)| \le C_n \left(\rho^{-n} \int_{B(a,\rho)} |u(w)|\, dw + \int_{B(a,\rho)} |\nabla u(w)|\, |w-a|^{1-n}\, dw \right).$$

This holds initially for smooth u. It also holds for a continuous $u \in H^{1,p}_{\mathrm{loc}}(\mathbb{R}^n)$ in the cases considered here: mollify u near $\overline{B(a,\rho)}$ and pass to the limit. The values at a converge by continuity, while the derivatives converge in L^p. For $p > n$, the weight $|w-a|^{1-n}$ belongs to $L^{p'}(B(a,\rho))$; when $p = n = 1$, the weight is 1.

Let $g \in H^{1,p}_{\mathrm{loc}}(\mathbb{R}^n)$ be continuous, where either $p > n$ or $p = n = 1$. Fix $x \in \mathbb{R}^n$ and $r > 0$, and write $B = B(x, 2r)$ and

$$g_B = \fint_B g(w)\, dw.$$

For any $z \in B(x,r)$ we have $B(z,r) \subseteq B$. Apply (78) to $u = g - g_B$, centered at z with radius r. This gives

$$|g(z) - g_B| \le C_n r^{-n} \int_{B(z,r)} |g(w) - g_B|\, dw + C_n \int_{B(z,r)} |\nabla g(w)|\, |w-z|^{1-n}\, dw. \tag{82}$$

We estimate the two terms separately. The L^1 Poincaré inequality on the ball B gives

$$\int_B |g - g_B|\, dw \le C_n r \int_B |\nabla g|\, dw. \tag{83}$$

Two proofs of (83). The fundamental theorem of calculus along the segment from a to b gives

$$g(a) - g(b) = \int_0^1 \nabla g\big((1-t)a + tb\big) \cdot (a-b)\, dt.$$

Since $a, b \in B(x,2r)$, we have $|a-b| \le 4r$. Consequently,

$$|g(a) - g(b)| \le 4r \int_0^1 \big|\nabla g\big((1-t)a + tb\big)\big|\, dt. \tag{84}$$

For $0 \le t \le \frac{1}{2}$, fix b and put $y = (1-t)a + tb$. Then $da = (1-t)^{-n} dy$. Convexity of B gives $(1-t)B + tb \subseteq B$, and $(1-t)^{-n} \le 2^n$. Hence

$$\begin{aligned} \int_B \int_B \big|\nabla g\big((1-t)a + tb\big)\big|\, da\, db &= \int_B (1-t)^{-n} \int_{(1-t)B+tb} |\nabla g(y)|\, dy\, db \\ &\le 2^n |B| \int_B |\nabla g(y)|\, dy. \end{aligned} \tag{85}$$

For $\frac{1}{2} \le t \le 1$, fix a instead and change variables in b. Now $db = t^{-n} dy$, $t^{-n} \le 2^n$, and $(1-t)a + tB \subseteq B$. Thus the same bound holds throughout $0 \le t \le 1$. Integrating in t and using (85) and (84), we obtain

$$\int_B |g - g_B| \le 4r\, 2^n \int_B |\nabla g| = 2^{n+2} r \int_B |\nabla g|.$$

Dividing by $|B|$ proves the claimed inequality.

A variational proof. The underlying L^1 Poincaré inequality can also be obtained by showing that

$$\inf\left\{\int_B |\nabla v| : v \in W^{1,1}(B), \quad v_B = 0, \quad \int_B |v| = 1\right\} > 0.$$

If this infimum were zero, there would be v_k with mean zero, L^1 norm one, and $\int_B |\nabla v_k| \to 0$. The sequence is bounded in $W^{1,1}(B)$. Compactness of the embedding $W^{1,1}(B) \hookrightarrow L^1(B)$ gives[36], after passage to a subsequence, $v_k \to v$ in $L^1(B)$. The weak gradient of v is zero[37], so v is constant on B. Its mean is zero, hence $v = 0$, contradicting $\|v\|_{L^1(B)} = 1$.

This proves the inequality on the unit ball. Rescaling gives

$$\int_B |g - g_B| \le C_n r \int_B |\nabla g|.$$

□

Consequently, Hölder's inequality gives

$$\begin{aligned} r^{-n}\int_{B(z,r)} |g - g_B|\, dw &\le r^{-n}\int_B |g - g_B|\, dw \\ &\le C_n r^{1-n}\int_B |\nabla g|\, dw \\ &\le C_{n,p} r^{1-n/p}\|\nabla g\|_{L^p(B)}. \end{aligned}$$

For the second term in (82), suppose first that $p > n$ and put $p' = p/(p-1)$. Then

$$\begin{aligned} \int_{B(z,r)} |\nabla g(w)|\, |w - z|^{1-n}\, dw &\le \|\nabla g\|_{L^p(B)} \left(\int_{B(z,r)} |w - z|^{(1-n)p'}\, dw\right)^{1/p'} \\ &= C_{n,p} r^{1-n/p}\|\nabla g\|_{L^p(B)}. \end{aligned}$$

---

[36] Let $\Omega \subset \mathbb{R}^n$ be a bounded Lipschitz domain and let $1 \le p \le \infty$. If $p < n$, then $H^{1,p}(\Omega) \hookrightarrow L^q(\Omega)$ compactly for every $1 \le q < p^* := np/(n-p)$. If $p = n$, the embedding is compact for every $1 \le q < \infty$. If $p > n$, it is compact into $C(\overline{\Omega})$, hence also into $L^q(\Omega)$ for every $1 \le q \le \infty$. More precisely, when $n < p < \infty$, it is compact into $C^{0,\alpha}(\overline{\Omega})$ for every $0 < \alpha < 1 - n/p$; when $p = \infty$, this holds for every $0 < \alpha < 1$. For $p < n$, the embedding into $L^{p^*}(\Omega)$ is continuous but generally not compact.

[37] Suppose that $v_k \to v$ in $L^1(B)$ and

$$\|\nabla v_k\|_{L^1(B)} \longrightarrow 0.$$

Fix $j \in \{1, \dots, n\}$ and $\varphi \in C_c^\infty(B)$. Since $\partial_j \varphi$ is bounded,

$$\int_B v_k\, \partial_j \varphi\, dx \longrightarrow \int_B v\, \partial_j \varphi\, dx.$$

On the other hand, the definition of the weak derivative gives

$$\int_B v_k\, \partial_j \varphi\, dx = -\int_B (\partial_j v_k)\varphi\, dx,$$

and

$$\left|\int_B (\partial_j v_k)\varphi\, dx\right| \le \|\varphi\|_{L^\infty(B)}\|\partial_j v_k\|_{L^1(B)} \longrightarrow 0.$$

Consequently,

$$\int_B v\, \partial_j \varphi\, dx = 0 \qquad \text{for every } \varphi \in C_c^\infty(B).$$

By definition, this says that $\partial_j v = 0$ weakly. Since this holds for every j, the weak gradient of v is zero. Because B is connected, v is constant almost everywhere on B.

The integral in the first line is finite precisely because $p > n$. If $p = n = 1$, its weight is 1, and the same estimate reads

$$\int_{B(z,r)} |\nabla g(w)|\, dw \le \|\nabla g\|_{L^1(B)}.$$

Thus (82) implies, uniformly for $z \in B(x,r)$,

$$|g(z) - g_B| \le C_{n,p} r^{1-n/p} \|\nabla g\|_{L^p(B(x,2r))}.$$

Applying this bound at two points of $B(x,r)$ yields

$$\sup_{y,z \in B(x,r)} |g(y) - g(z)| \le C_{n,p} r^{1-n/p} \|\nabla g\|_{L^p(B(x,2r))}. \tag{86}$$

Notice that this is a local statement: g need only belong to $H^{1,p}$ near $B(x,2r)$.

Now take the continuous representative of f, and let x be an L^p Lebesgue point of ∇f. Set $v = \nabla f(x)$ and

$$g(y) = f(y) - v \cdot y.$$

Although the affine function $v \cdot y$ need not belong to $H^{1,p}(\mathbb{R}^n)$, we have $g \in H^{1,p}_{\mathrm{loc}}(\mathbb{R}^n)$ and $\nabla g = \nabla f - v$ almost everywhere. Since $x \in B(x,r)$, (86) gives, for every $y \in B(x,r)$,

$$|f(y) - f(x) - v \cdot (y-x)| \le C_{n,p} r^{1-n/p} \left(\int_{B(x,2r)} |\nabla f(z) - v|^p \, dz \right)^{1/p}.$$

Using $|B(x,2r)| = \omega_n (2r)^n$, the right-hand side is

$$C_{n,p} r \left(\fint_{B(x,2r)} |\nabla f(z) - \nabla f(x)|^p \, dz \right)^{1/p} = o(r),$$

because x is an L^p Lebesgue point of ∇f.

Finally, given $y \neq x$, choose $r = 2|y-x|$. Then $y \in B(x,r)$, and the preceding estimate gives

$$\frac{|f(y) - f(x) - \nabla f(x) \cdot (y-x)|}{|y-x|} \longrightarrow 0 \qquad \text{as } y \to x.$$

Hence f is differentiable at x, with $Df(x) = \nabla f(x)$. □

Corollary 9.17 (Rademacher's theorem)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open and let $f : \Omega \to \mathbb{R}^m$ be Lipschitz. Then f is differentiable almost everywhere in Ω.*

Proof. Choose relatively compact open sets Ω_k such that $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$, and choose $\varphi_k \in C_c^\infty(\Omega)$ with $\varphi_k = 1$ on Ω_k. Extend $\varphi_k f$ by zero outside Ω. This extension is bounded, Lipschitz, and compactly supported. By the preceding proposition, it belongs to $H^{1,\infty}(\mathbb{R}^n)$; compact support then gives $\varphi_k f \in H^{1,p}(\mathbb{R}^n)$ for every finite $p \ge 1$.

Fix $p > n$. By Theorem 9.16, each component of $\varphi_k f$ is differentiable almost everywhere. Since $\varphi_k f = f$ on Ω_k, the same holds for f on Ω_k. The union of the corresponding exceptional sets is null, so f is differentiable almost everywhere in Ω. □

10. HAUSDORFF MEASURES

Let (X,d) be a separable metric space, and let

$$h^*_{r,\delta}(S) = \inf \left\{ \sum_{j \ge 1} (\operatorname{diam} S_j)^r : S \subseteq \bigcup_{j \ge 1} S_j, \operatorname{diam} S_j \le \delta \right\}, \tag{87}$$

where $S_j \subseteq X$, with the convention that $\inf \emptyset = \infty$. It is immediate that $h^*_{r,\delta}$ is an outer measure, and monotonically decreasing in δ. We define

$$h^*_r(S) = \lim_{\delta \to 0} h^*_{r,\delta}(S). \tag{88}$$

Again, we leave the following as an exercise:

Exercise 10.1. *h^*_r is a metric outer measure.*

For every h^*_r-measurable set S (which includes the Borel sets), we set

$$\mathcal{H}^r(S) = \sigma_r h^*_r(S), \text{ where } \sigma_r = \frac{\pi^{r/2} 2^{-r}}{\Gamma(\frac{r}{2}+1)}. \tag{89}$$

When $r \in \mathbb{N}$, σ_n is the volume of the Euclidean n-ball of diameter 1.

Let us make a few comments about how to calculate the volume of an Euclidean n-ball of radius 1. By integration in the radial direction[38], it suffices to find the surface area of the sphere S^{n-1}. We quickly indicate how to do that. One way is to go up inductively by knowing the surface areas of lower dimensional spheres. A slick method is to observe that the integral $\int_0^\infty \int_{S^{n-1}} e^{-r^2} r^{n-1}\, dr d\omega$ can be written as $\prod_{j=1}^n \int_0^\infty e^{-x_j^2}\, dx_j$. However, the first integral can be expressed in terms of the Gamma function, like so:

$$\begin{aligned} \int_0^\infty e^{-r^2} r^{n-1}\, dr d\omega &= \frac{1}{2} \left|S^{n-1}\right| \int_0^\infty e^{-z} z^{\frac{n}{2}-1}\, dz \\ &= \frac{1}{2} \left|S^{n-1}\right| \Gamma\left(\frac{n}{2}\right). \end{aligned}$$

Also, observe that each of $\int_0^\infty e^{-x_j^2}\, dx_j$ can be expressed as $\frac{1}{2}\Gamma\left(\frac{1}{2}\right)$. We leave it to the reader to finish the computation.

Let $|.|$ denote the Lebesgue measure on $\mathbb{R}^n$. The main reason for the scaling with ω_n is to show that $\mathcal{H}^n = |.|$. We first prove the following proposition:

Proposition 10.2 (Isodiametric inequality)**.** *For every Lebesgue measurable set $S \subseteq \mathbb{R}^n$,*

$$|S| \leq \sigma_n (\operatorname{diam} S)^n, \tag{90}$$

where $\sigma_n = \omega_n / 2^n$ is the volume of the ball of diameter 1; we take $\operatorname{diam} \emptyset = 0$.

We give a proof by Steiner symmetrisation.[39]

Proof. We first consider a nonempty compact set $K \subseteq \mathbb{R}^n$. Let η be a unit vector, let $P_\eta = \eta^\perp$, and let π_η denote orthogonal projection onto P_η. For $y \in P_\eta$, put

$$\ell_K(y) := \int_{\mathbb{R}} \chi_K(y + t\eta)\, dt, \qquad Y := \pi_\eta(K).$$

The *Steiner symmetral* of K in the direction η is

$$\Sigma_\eta(K) := \left\{ y + t\eta : y \in Y,\ |t| \leq \tfrac{1}{2} \ell_K(y) \right\}. \tag{91}$$

Thus each nonempty section parallel to η is replaced by a centered interval of the same length. Sections of length zero give single points.

[38]The integration formula in polar coordinates will be a topic of presentation. In any case, such a formula can be derived easily from the co-area formula (see below).

[39]See A. Treibergs, *Steiner Symmetrization and Applications*, slides 7–11, `http://www.math.utah.edu/~treiberg/Steiner/SteinerSlides.pdf`.

Let us check that $\Sigma_\eta(K)$ is compact. The set Y is compact. Also, ℓ_K is upper semicontinuous. Indeed, if $y_j \to y$, choose R so that $K \subseteq B(0,R)$. Since K is closed, for each t we have

$$\limsup_{j\to\infty} \chi_K(y_j + t\eta) \le \chi_K(y + t\eta).$$

All these functions vanish when $|t| > R$. Applying reverse Fatou on $[-R, R]$ gives

$$\limsup_{j\to\infty} \ell_K(y_j) \le \ell_K(y).$$

It follows that the set in (91) is closed and bounded, hence compact. Fubini's theorem, in the orthogonal coordinates $x = y + t\eta$, now gives

$$|\Sigma_\eta(K)| = \int_{P_\eta} \ell_K(y)\, dy = |K|. \tag{92}$$

Here dy is Lebesgue measure on P_η; when $n = 1$, it is counting measure on the one-point space P_η.

We next prove that

$$\operatorname{diam} \Sigma_\eta(K) \le \operatorname{diam} K. \tag{93}$$

Write $D = \operatorname{diam} K$, and take $y, z \in Y$. The nonempty compact sections

$$I_y := \{t : y + t\eta \in K\}, \qquad I_z := \{t : z + t\eta \in K\}$$

have minima a_y, a_z and maxima b_y, b_z, respectively. Put $a = \ell_K(y)$ and $b = \ell_K(z)$. Since a measurable subset of an interval has length at most that of the interval,

$$a \le b_y - a_y, \qquad b \le b_z - a_z.$$

Consequently,

$$d := \max\{b_y - a_z,\ b_z - a_y\} \ge \frac{(b_y - a_y) + (b_z - a_z)}{2} \ge \frac{a+b}{2}.$$

The endpoint pairs defining d are pairs of points of K. Their orthogonal components differ by $y - z$, so

$$D^2 \ge |y - z|^2 + d^2.$$

On the other hand, if $y + t\eta, z + s\eta \in \Sigma_\eta(K)$, then $|t - s| \le (a+b)/2 \le d$. Hence

$$|(y + t\eta) - (z + s\eta)|^2 = |y - z|^2 + |t - s|^2 \le D^2,$$

which proves (93).

Choose an orthonormal basis $e_1, \dots, e_n$, and define

$$K_0 = K, \qquad K_j = \Sigma_{e_j}(K_{j-1}) \quad (1 \le j \le n).$$

Every K_j is compact and, by (92) and (93), satisfies $|K_j| = |K|$ and $\operatorname{diam} K_j \le D$. Let R_i denote reflection in $e_i^\perp$. The set K_j is invariant under R_j by construction. Moreover, if $i \ne j$, reflection by R_i maps lines parallel to e_j onto lines parallel to e_j and preserves the lengths of their sections. Thus $R_i K_{j-1} = K_{j-1}$ implies $R_i K_j = K_j$. By induction, K_n is invariant under all the R_i, and therefore under $R_1 \cdots R_n = -\mathrm{id}$. For every $x \in K_n$ we have $-x \in K_n$; hence $2|x| \le \operatorname{diam} K_n \le D$. We conclude that

$$|K| = |K_n| \le |B(0, D/2)| = \omega_n (D/2)^n = \sigma_n D^n.$$

Finally, let S be Lebesgue measurable. If $|S| = 0$ or $\operatorname{diam} S = \infty$, the assertion is immediate. Otherwise S is bounded and $0 < |S| < \infty$. By inner regularity, there are compact sets $K_j \subseteq S$ with $|K_j| \to |S|$. Since $\operatorname{diam} K_j \le \operatorname{diam} S$, applying the compact case and passing to the limit proves the inequality for S. □

Remark 10.3. *The isodiametric inequality would be trivially true if every set is contained in a ball of the same diameter. But this fails even for equilateral triangles inscribed in a circle.*

The following exercise gives additional information about the symmetrisation step:

Exercise 10.4. *Let $\gamma : [0,a] \to \mathbb{R}^n$ be a continuous injective function. The length of γ is defined as $l(\gamma) = \sup\{\sum_{j=1}^{N} |\gamma(t_j) - \gamma(t_{j-1})| : 0 = t_0 < \cdots < t_N\}$. Then prove that*

$$l(\gamma) = \mathcal{H}^1(\gamma([0,a])).$$

The following is left as an exercise:

Exercise 10.5. *For any measurable set A, we have that $|A| \leq \mathcal{H}^n(A)$. (Hint: Apply the isodiametric inequality and take infimum).*

We now prove the other side:

Theorem 10.6. *If $U \subseteq \mathbb{R}^n$ is open, then $\mathcal{H}^n(U) \leq |U|$.*

Proof. Assume that $|U| < \infty$, and write $U = \bigcup_\alpha B_\alpha^\delta$, where each ball has diameter $\leq \delta$. By Wiener's covering lemma, one has a finite subcollection, still called $B_j, 1 \leq j \leq K$ with total Lebesgue measure $> 4^{-n}|U|$. Apply this argument again to $U \setminus \bigcup_{j=1}^{K} \overline{B_j}$ and repeat the process countably many times, we get a countable collection B_k such that $|U \setminus \bigcup_{k\geq 1} B_k| = 0$, implying that $\omega_n h^*_{n,\delta}(U_1) \leq |U|$, where $U_1 := \bigcup_{k\geq 1} B_k$.

Now, we repeat the process with U replaced by U_1 and δ replaced by $\delta/2$. The lth step will continue with δ replaced by $\delta/2^l$ and we shall obtain an open set U_l. Now, define

$$V = \bigcap_{l\geq 1} U_l.$$

Clearly, $\mathcal{H}^n(V) \leq |U|$. On the other hand,

$$|U \setminus V| = 0 \implies \mathcal{H}^n(U \setminus V) = 0 \implies \mathcal{H}^n(U) = \mathcal{H}^n(V).$$

□

Exercise 10.7. *Justify the last step in the above proof by proving the following:*

$$\mathcal{H}^n(U) \leq \omega_n n^{n/2} |U|.$$

(Hint: Cover U by a countable family of cubes of side length ρ whose volumes sum to $\leq |U| + \varepsilon$. Since diameter of each cube is $\rho\sqrt{n}$....)

We now prove the following

Theorem 10.8. *For any Borel set $A \subseteq \mathbb{R}^n$,*

$$|A| = \mathcal{H}^n(A). \tag{94}$$

It suffices to prove this for compact sets, if we are able to establish inner regularity for the Hausdorff measure. Let K be a compact set, and let B be an open ball containing K. But, we have that $\mathcal{H}^n(B) = |B|$, and by the above theorem, and $\mathcal{H}^n(B \setminus K) = |B \setminus K|$. Then, $\mathcal{H}^n(K) = |K|$. The bit about the inner regularity is left as an exercise:

Exercise 10.9. *Prove that the Hausdorff measure is Borel inner regular, that is, for any Borel set A, we have that $\mathcal{H}^n(A) = \sup\{\mathcal{H}^n(K) : K \subseteq A, K \text{ compact }\}$.*

This is not completely easy as far as I can tell. One needs to repeat an argument which is somewhat similar in spirit to Theorem 10.6. Have a look at [Fal86, Theorem 1.6(b)] if you wish. Another way to do this is via [Fol99, Theorem 7.8], which implies that on $\mathbb{R}^n$ every Borel measure that is finite on compact sets is already Radon.

Remark 10.10. *[Fol99, Proposition 11.20] gives a different proof of Theorem 10.8 above. Basically, he just checks that $\mathcal{H}^n$ is a left Haar measure*[40]*, so it has to be (up to a constant) the Lebesgue measure. However, we took a more hands-on approach as we have not proved the last stated fact.*

Now, we discuss a bit about $\mathcal{H}^r$, where r is not an integer. We first define the concept of Hausdorff dimensions, as follows: let S be a non-empty subset of a metric space X. Then,

$$\operatorname{Hdim} S = \inf\{r \geq 0 : \mathcal{H}^r(S) = 0\} = \sup\{r \geq 0 : \mathcal{H}^r(S) > 0\}. \tag{95}$$

Observe that the following lemma is almost immediate:

Lemma 10.11. *Let $S \subseteq \mathbb{R}^n$, and let $k = \operatorname{Hdim} S$. If $r < k$, then $\mathcal{H}^r(S) = \infty$. If $r > k$, then $\mathcal{H}^r(S) = 0$.*

Proof. One can calculate that

$$\begin{aligned} h^*_{r,\delta}(S) &= \inf\left\{\sum_j (\operatorname{diam} S_j)^r : \operatorname{diam} S_j \leq \delta, S \subseteq \cup S_j\right\} \\ &= \inf\left\{\sum_j (\operatorname{diam} S_j)^k (\operatorname{diam} S_j)^{r-k} : \operatorname{diam} S_j \leq \delta, S \subseteq \cup S_j\right\} \\ &\leq \delta^{r-k} \inf\left\{\sum_j (\operatorname{diam} S_j)^k : \operatorname{diam} S_j \leq \delta, S \subseteq \cup S_j\right\}. \end{aligned}$$

If $r > k$, then it is clear that $h^*_{r,\delta}(S) \leq \delta^{r-k} h^*_{k,\delta}(S)$, which implies that $\mathcal{H}^r(S) = 0$.

By a similar calculation, in the case $r < k$ one can see that

$$h^*_{k,\delta}(S) \leq \delta^{k-r} h^*_{r,\delta}(S),$$

which implies that if $r < k$, then $\mathcal{H}^r(S) = \infty$. □

An interesting example is the Cantor middle third set K[41].

Let K_j be the usual sequence such that $K_j \searrow K$. Each K_j is a cover of K by 2^j intervals of length 3^{-j}, so $h^*_{r,\delta}(K) \leq \left(\frac{2}{3^r}\right)^j, \delta = 3^{-j}$. The RHS is independent of j if $3^r = 2$ or if $r = \log 2/\log 3$. The second part of the above lemma already implies that $\operatorname{Hdim} K \leq \log 2/\log 3$. We leave the following as a (non-trivial) exercise:

Exercise 10.12. $\operatorname{Hdim} K = \frac{\log 2}{\log 3}$.

Parallel to the construction of the Cantor set is the following 2-dimensional analogue: consider an equilateral triangle in the plane, and subdivide it into 4 similar equilateral triangles by forming a triangle with the midpoints of the three sides. Now, delete the middle triangle. In each of the three remaining triangles, repeat the same procedure, and continue inductively. The compact set obtained by taking the infinite intersection is called the Sierpinski triangle. It can be proved that the Sierpinski triangle has Hausdorff dimension $\log 3/\log 2$.

We now state the following efficient criteria for estimating the Hausdorff dimension of a set:

Theorem 10.13. *Let $K \subseteq \mathbb{R}^n$ be compact, and $a \in (0, \infty)$.*

[40] It suffices to check left translation invariance (which is obvious), finiteness on compact sets (again obvious) and Radon (the latter has been indicated in the Exercise above).

[41] If it is not already known, convince yourself that $|K| = 0$.

(1) *If there exists a positive Borel measure supported on K such that*

$$\int\int \frac{1}{|x-y|^a}\, d\mu(x)d\mu(y) < \infty, \tag{96}$$

then $\mathcal{H}^a(K) = \infty$, which implies that $\operatorname{Hdim} K \geq a$.

(2) *If for each non-zero positive Borel measure μ supported on K one has*

$$\int\int \frac{1}{|x-y|^a}\, d\mu(x)d\mu(y) = \infty,$$

then $\mathcal{H}^b(K) = 0$ for all $b > a$. This in turn implies that $\operatorname{Hdim} K \leq a$.

We only prove the first part. The second part can be found in [Fal86].

Proof. Assume that $\mathcal{H}^a(K) < \infty$ and we shall show that the integral is infinite. Consider the set

$$E = \{x \in K : \limsup_{r\to 0} r^{-a}\mu(B_r(x)) > 0\}. \tag{97}$$

First, we prove that $\mu(K \setminus E) = 0$. Given $\varepsilon, \rho > 0$, set

$$F_{\varepsilon,\rho} = \{x \in K : \mu(B_r(x)) \leq \varepsilon r^a,\ \forall r \leq \rho\}.$$

Take a countable cover U_j of $F_{\varepsilon,\rho}$ by sets of diameter $\leq \delta \leq \rho$, and take balls B_j centred in $F_{\varepsilon,\rho}$ such that $B_j \supseteq U_j$ and $\operatorname{diam} B_j \leq 2 \operatorname{diam} U_j \leq 2\rho$[42]. Then,

$$\mu(F_{\varepsilon,\rho}) \leq \sum_j \mu(B_j) \leq 2^{-a}\varepsilon \sum_j (\operatorname{diam} B_j)^a \leq \varepsilon \sum_j (\operatorname{diam} U_j)^a.$$

Hence, $\mu(F_{\varepsilon,\rho}) \leq \varepsilon \mathcal{H}^a_\delta(F_{\varepsilon,\rho})$, and the latter (after taking $\delta \to 0$) is $\leq \varepsilon \mathcal{H}^a(K)$ $\forall \varepsilon, \rho > 0$. Since $E \supseteq K \setminus \bigcap_{\varepsilon,\rho} F_{\varepsilon,\rho}$, this proves that $\mu(K \setminus E) = 0$.

If $x \in E$, then there is a sequence $r_j \searrow 0$ such that $\mu(B_{r_j}(x)) \geq C r_j^a$ for some $C > 0$. We can assume that $\mu(\{x\}) = 0$, since otherwise the integral in (96) is clearly infinite. Then there exists $q_j \in (0, r_j)$ such that

$$\mu(A_j) \geq \frac{1}{2} C r_j^a, \quad A_j = B_{r_j}(x) \setminus B_{q_j}(x).$$

Passing to a subsequence to ensure that $r_{j+1} < q_j$, the shells A_j are disjoint. Then,

$$\begin{aligned} x \in E \implies \int_K \frac{d\mu(y)}{|x-y|^a} &\geq \sum_j \int_{A_j} \frac{d\mu(y)}{|x-y|^a} \\ &\geq \frac{C}{2} \sum_{j\geq 1} r_j^a r_j^{-a} \\ &= +\infty. \end{aligned}$$

Since $\mu(K \setminus E) = 0$, we have that $\mu(E) > 0$, which in turn implies that the integral in (96) is

$$\geq \int_E \int_K \frac{d\mu(y)\, d\mu(x)}{|x-y|^a} = \infty.$$

□

In light of the above, we make the following definition:

[42] Discard any U_j that does not meet $F_{\varepsilon,\rho}$. Otherwise choose $x_j \in U_j \cap F_{\varepsilon,\rho}$ and put $d_j = \operatorname{diam} U_j$. The closed ball $B_j = \overline{B}(x_j, d_j)$ contains U_j, is centred in $F_{\varepsilon,\rho}$, and satisfies $\operatorname{diam} B_j \leq 2d_j$.

Definition 10.14. *Let μ be a finite measure on a metric space X and $\alpha \geq 0$. Then the α-potential of a point $x \in X$ with respect to μ is defined by $P_\alpha(x) := \int \frac{d\mu(y)}{d(x,y)^\alpha}$. The α-energy of μ is $I_\alpha(\mu) = \int P_\alpha \, d\mu(x) = \int\int \frac{d\mu(x)d\mu(y)}{d(x,y)^\alpha}$.*

The α-capacity of a set E is defined to be

$$\operatorname{Cap}_\alpha(E) := \sup\{I_\alpha(\mu)^{-1} : \mu \text{ is a Borel probability measure supported on } E\}.$$

Theorem 10.15. *Let $E \subseteq \mathbb{R}^n$ be a nonempty closed set. Then*

$$\operatorname{Hdim} E = \sup\{\alpha \geq 0 : \operatorname{Cap}_\alpha(E) > 0\}.$$

Here the kernel for $\alpha = 0$ is understood to be identically 1, so that $\operatorname{Cap}_0(E) = 1$.

Proof. We use the following form of Frostman's lemma.

Lemma 10.16 (Frostman's lemma)**.** *Let $K \subseteq \mathbb{R}^n$ be compact and let $s > 0$. Then $\mathcal{H}^s(K) > 0$ if and only if there exist a Borel probability measure μ supported on K and a constant $C < \infty$ such that*

$$\mu(B(x,r)) \leq Cr^s \qquad \text{for every } x \in \mathbb{R}^n \text{ and } r > 0.$$

Suppose first that $\alpha > 0$ and $\operatorname{Cap}_\alpha(E) > 0$. Then there is a Borel probability measure μ supported on E with $I_\alpha(\mu) < \infty$. Choose $R > 0$ such that $\mu(K) > 0$, where $K = E \cap \overline{B(0,R)}$. The normalized restriction

$$\nu = \frac{\mu|_K}{\mu(K)}$$

is supported on the compact set K, and

$$I_\alpha(\nu) \leq \frac{I_\alpha(\mu)}{\mu(K)^2} < \infty.$$

By the first part of Theorem 10.13, $\operatorname{Hdim} K \geq \alpha$. Thus $\operatorname{Hdim} E \geq \alpha$.

Conversely, fix $0 < \alpha < \operatorname{Hdim} E$ and choose $\alpha < s < \operatorname{Hdim} E$. We have $\mathcal{H}^s(E) > 0$. Since

$$E = \bigcup_{m=1}^{\infty} \big(E \cap \overline{B(0,m)}\big),$$

countable subadditivity shows that, for some m, the compact set $K = E \cap \overline{B(0,m)}$ satisfies $\mathcal{H}^s(K) > 0$. Frostman's lemma gives a Borel probability measure μ supported on K such that

$$\mu(B(x,r)) \leq Cr^s \qquad (x \in \mathbb{R}^n,\ r > 0).$$

For each $x \in \mathbb{R}^n$, Tonelli's theorem gives

$$\begin{aligned}\int_{\mathbb{R}^n} |x-y|^{-\alpha}\, d\mu(y) &= \alpha \int_0^\infty r^{-\alpha-1}\mu(B(x,r))\, dr \\ &\leq C\alpha \int_0^1 r^{s-\alpha-1}\, dr + \alpha \int_1^\infty r^{-\alpha-1}\, dr \\ &= \frac{C\alpha}{s-\alpha} + 1 < \infty.\end{aligned}$$

The bound is independent of x. Integrating with respect to $\mu(x)$ shows that $I_\alpha(\mu) < \infty$, and hence $\operatorname{Cap}_\alpha(E) > 0$.

We have proved that every positive $\alpha < \operatorname{Hdim} E$ has positive α-capacity, whereas positive α-capacity implies $\alpha \leq \operatorname{Hdim} E$. The convention at $\alpha = 0$ covers the case $\operatorname{Hdim} E = 0$. □

Remark 10.17. *We have already mentioned in class the concept of capacity that is tied to Sobolev spaces. The above definition gives another notion of capacity. The importance of these results lies in the fact that no matter what the notion of capacity is, in most cases it is comparable to hitting probabilities of Brownian particles. In general, capacities are difficult to estimate directly, but there is a well-developed literature about hitting probabilities and their estimates. We shall not have time to delve into this matter any further in this course, as this gets into pretty deep waters pretty quickly. However, as a sample, I recommend the reader to have a look at [MP10, Theorem 8.24], which discusses the related concept of Martin capacity. We shall have the opportunity to cover Brownian motion and Wiener measures later on in our course, so that will provide added perspective on the foregoing remark.*

Remark 10.18. *The topological dimension of a topological space X is defined to be the minimum value of m such that any open cover has a refinement with multiplicity $\leq n+1$. Following Mandelbrot, a set S is called a fractal if its topological dimension disagrees with its Hausdorff dimension. The Cantor set is the first basic example of a fractal. Brownian paths in $\mathbb{R}^n, n \geq 2$ which we shall see later on are also fractals. See [Fal86] for more details.*

10.1. **Isoperimetric inequalities.** According to [Sim83], a Borel set $E \subseteq \mathbb{R}^n$ is called m-rectifiable if E is of Hausdorff dimension m and there exist a countable collection $f_j : \mathbb{R}^m \to \mathbb{R}^n$ of Lipschitz maps such that $\mathcal{H}^m\left(E \setminus \bigcup_{j\geq 1} f_j(\mathbb{R}^m)\right) = 0$. An alternative definition is to require that E is a countable union of sets which are the image of Lipschitz maps from some bounded subset of $\mathbb{R}^m$.

Let E be a subset of a metric space X. Then, E is called m-rectifiable if there exists a Lipschitz map from some bounded subset of $\mathbb{R}^m$ onto E. E is called countably m-rectifiable if E equals the union of some countable family whose members are m-rectifiable. The latter definition is from [Fed69, 3.2.14]. I am not sure of the equivalence.

The following is a prototypical isoperimetric inequality, whose proof will be the topic of a presentation[43]:

Theorem 10.19. *Let $\Omega \subseteq \mathbb{R}^n$ be a domain whose boundary $\partial\Omega$ is $(n-1)$-rectifiable. Then, we have the sharp inequality*

$$|\Omega|^{\frac{n-1}{n}} \leq \frac{1}{n\omega_n^{1/n}} \mathcal{H}^{n-1}(\partial\Omega). \tag{98}$$

10.2. **Co-area formula.** Let $N \geq n \geq 1$, and put $k = N - n$. We normalize $\mathcal{H}^k$ so that it agrees with k-dimensional Lebesgue measure on every k-plane. In particular, $\mathcal{H}^0$ is counting measure. For a Lipschitz map $F : \mathbb{R}^N \to \mathbb{R}^n$, Rademacher's theorem gives DF almost everywhere. At such points set

$$JF(x) := \sqrt{\det\big(DF(x)DF(x)^{\mathsf{T}}\big)}.$$

Its value on the remaining null set is immaterial; we set it equal to zero. We write $\mathcal{H}^k_\delta(A)$ for the usual Hausdorff covering quantity obtained by restricting the diameters of the covering sets to be less than δ. Thus $\mathcal{H}^k_\delta(A)$ increases to $\mathcal{H}^k(A)$ as $\delta \downarrow 0$. When $k = 0$, each nonempty member of a cover has weight one.

Theorem 10.20 (Co-area formula)**.** *For every Borel set $E \subseteq \mathbb{R}^N$,*

$$\int_E JF(x)\, dx = \int_{\mathbb{R}^n} \mathcal{H}^k\big(E \cap F^{-1}(y)\big)\, dy. \tag{99}$$

[43]More formally, the topic of the presentation will be the equivalence between the isoperimetric inequality and a sharp Sobolev inequality.

Consequently, for every nonnegative[44] *Lebesgue measurable* $g : \mathbb{R}^N \to [0, \infty]$,

$$\int_{\mathbb{R}^N} g(x) JF(x)\, dx = \int_{\mathbb{R}^n} \left(\int_{F^{-1}(y)} g(x)\, d\mathcal{H}^k(x) \right) dy. \tag{100}$$

The formula for sets also holds when E *is Lebesgue measurable. For a Lebesgue measurable* g*, the inner integral is understood for almost every* y*, as established below.*

Proof. **1. Projection is Fubini's theorem**

Write $x = (u, v) \in \mathbb{R}^n \times \mathbb{R}^k$, and consider $P(u, v) = u$. Then $JP = 1$, while $P^{-1}(y) = \{y\} \times \mathbb{R}^k$. For this map, (99) reads

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \mathbf{1}_E(u, v)\, dv\, du = \int_{\mathbb{R}^n} \mathcal{H}^k\big(E \cap (\{y\} \times \mathbb{R}^k)\big)\, dy.$$

This is Fubini's theorem: the $\mathcal{H}^k$-measure of the slice on the right is its dv-measure.

2. A surjective linear map

First suppose that $L(u, v) = Mu$, where M is an invertible $n \times n$ matrix. Since $LL^{\mathsf{T}} = MM^{\mathsf{T}}$, we have $JL = |\det M|$. Thus the left-hand side of (99) is

$$|\det M| \int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \mathbf{1}_E(u, v)\, dv\, du.$$

For fixed y, the fibre $L^{-1}(y)$ consists of the points $(M^{-1}y, v)$. The right-hand side is therefore

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^k} \mathbf{1}_E(M^{-1}y, v)\, dv\, dy.$$

The change of variables $y = Mu$ makes the two expressions equal.

Every surjective linear map $L : \mathbb{R}^N \to \mathbb{R}^n$ reduces to this case. Choose orthogonal coordinates corresponding to

$$\mathbb{R}^N = (\ker L)^\perp \oplus \ker L.$$

The restriction of L to $(\ker L)^\perp$ is invertible, and orthogonal changes of coordinates preserve both Lebesgue and Hausdorff measure.

3. The area of a graph

We shall use the following formula for the Hausdorff measure of a C^1 graph.

Lemma 10.21 (Graph area formula)**.** *Let* $V \subseteq \mathbb{R}^k$ *be open, let* $h : V \to \mathbb{R}^n$ *be* C^1*, and define*

$$\gamma : V \longrightarrow \mathbb{R}^{n+k}, \qquad \gamma(v) = (h(v), v).$$

Then, for every Borel set $S \subseteq V$*,*

$$\mathcal{H}^k(\gamma(S)) = \int_S \sqrt{\det\big(I_k + Dh(v)^{\mathsf{T}} Dh(v)\big)}\, dv. \tag{101}$$

When $k = 0$*, both sides are interpreted using counting measure on* $\mathbb{R}^0$ *and the convention that the determinant of the empty matrix is* 1*.*

Idea. An affine graph has a fixed area factor, which we can calculate by linear change of variables. On a sufficiently small cube, a C^1 graph and its tangent affine graph distort distances by nearly the same amount. We apply the affine calculation on each small cube and then let their diameters tend to zero.

[44] Strictly speaking, we do not need nonnegative g. One of $\int_{\mathbb{R}^N} g^+(x) JF(x)\, dx$ or $\int_{\mathbb{R}^N} g^-(x) JF(x)\, dx$ need to be finite so that we are not in a $\infty - \infty$ situation.

Proof. If $k = 0$, the only nonempty subset of $\mathbb{R}^0$ consists of one point. Its image under γ also consists of one point, and both sides of (101) equal 1. We henceforth assume $k \geq 1$.

The affine case. Fix an $n \times k$ matrix C and a point $a \in \mathbb{R}^{n+k}$, and set

$$\gamma_C(v) = a + Av, \qquad A = \begin{pmatrix} C \\ I_k \end{pmatrix}.$$

The matrix A represents an injective linear map from $\mathbb{R}^k$ into $\mathbb{R}^{n+k}$. Indeed, the last k coordinates of Av are v. Consequently, $P = A(\mathbb{R}^k)$ is a k-dimensional linear subspace.

Choose an orthonormal basis of P. Taking coordinates in this basis defines a linear isometry $T : P \to \mathbb{R}^k$. Let M be the $k \times k$ matrix of the map $T \circ A : \mathbb{R}^k \to \mathbb{R}^k$. For $v, w \in \mathbb{R}^k$,

$$\langle Mv, Mw \rangle = \langle Av, Aw \rangle,$$

because T is an isometry. It follows that

$$M^{\mathsf{T}} M = A^{\mathsf{T}} A = \begin{pmatrix} C^{\mathsf{T}} & I_k \end{pmatrix} \begin{pmatrix} C \\ I_k \end{pmatrix} = I_k + C^{\mathsf{T}} C.$$

Taking determinants, we obtain

$$|\det M| = \sqrt{\det(I_k + C^{\mathsf{T}} C)}.$$

Let $S \subseteq \mathbb{R}^k$ be Borel. Translation by a and the isometry T preserve $\mathcal{H}^k$. Moreover, $\mathcal{H}^k$ on $\mathbb{R}^k$ is Lebesgue measure with our normalization. The ordinary linear change-of-variables formula therefore gives

$$\begin{aligned} \mathcal{H}^k(\gamma_C(S)) &= \mathcal{H}^k(A(S)) \\ &= \mathcal{H}^k(T(A(S))) \\ &= |M(S)| \\ &= |\det M|\,|S| \\ &= \sqrt{\det(I_k + C^{\mathsf{T}} C)}\,|S|. \end{aligned} \tag{102}$$

This is the formula for an affine graph.

Comparison with a tangent graph. Return to the given C^1 map h, and write

$$J_k\gamma(v) := \sqrt{\det\bigl(I_k + Dh(v)^{\mathsf{T}} Dh(v)\bigr)}.$$

Fix a closed cube $R \subseteq V$. We first prove (101) for every Borel set $S \subseteq R$.

Partition R into finitely many closed cubes Q_i of equal side length. Put

$$d := \max_i \operatorname{diam} Q_i$$

and choose $v_i \in Q_i$ for each i. The tangent affine graph at v_i is

$$\gamma_i(v) = \gamma(v_i) + D\gamma(v_i)(v - v_i), \qquad v \in \mathbb{R}^k.$$

Since $\gamma(v) = (h(v), v)$,

$$D\gamma(v_i) = \begin{pmatrix} Dh(v_i) \\ I_k \end{pmatrix}.$$

The affine formula (102) thus says that, for every Borel set $A \subseteq Q_i$,

$$\mathcal{H}^k(\gamma_i(A)) = J_k\gamma(v_i)\,|A|. \tag{103}$$

The derivative $D\gamma$ is uniformly continuous on R. Define

$$\varepsilon_d := \sup\bigl\{\|D\gamma(z) - D\gamma(z')\| : z, z' \in R,\ |z - z'| \leq d\bigr\}.$$

Then $\varepsilon_d \to 0$ as $d \to 0$. If $v, w \in Q_i$, the segment from w to v lies in Q_i. The fundamental theorem of calculus gives

$$\gamma(v) - \gamma(w) - D\gamma(v_i)(v-w) = \int_0^1 \big[D\gamma(w + t(v-w)) - D\gamma(v_i)\big](v-w)\,dt.$$

Every point $w + t(v-w)$ belongs to Q_i and is at distance at most d from v_i. Hence

$$|\gamma(v) - \gamma(w) - D\gamma(v_i)(v-w)| \le \varepsilon_d |v-w|. \tag{104}$$

The last k coordinates of $D\gamma(v_i)(v-w)$ are $v-w$, so

$$|\gamma_i(v) - \gamma_i(w)| = |D\gamma(v_i)(v-w)| \ge |v-w|.$$

Using this in (104) and applying the triangle inequality in both directions, we obtain

$$(1-\varepsilon_d)|\gamma_i(v) - \gamma_i(w)| \le |\gamma(v) - \gamma(w)| \le (1+\varepsilon_d)|\gamma_i(v) - \gamma_i(w)| \tag{105}$$

for all $v, w \in Q_i$. We take d small enough that $\varepsilon_d < 1$.

The map

$$\gamma \circ \gamma_i^{-1} : \gamma_i(Q_i) \longrightarrow \gamma(Q_i)$$

is therefore $(1+\varepsilon_d)$-Lipschitz, and its inverse is $(1-\varepsilon_d)^{-1}$-Lipschitz. A Lipschitz map with constant L increases $\mathcal{H}^k$ by at most a factor L^k. Applying this fact to the two maps, and then using (103), gives

$$(1-\varepsilon_d)^k J_k\gamma(v_i)|A| \le \mathcal{H}^k(\gamma(A)) \le (1+\varepsilon_d)^k J_k\gamma(v_i)|A| \tag{106}$$

for every Borel set $A \subseteq Q_i$.

Passing from small cubes to an arbitrary Borel set. If we wanted the formula only for the whole cube R, the remaining argument would be a Riemann-sum argument. But the lemma concerns every Borel set $S \subseteq R$. The function $\mathbf{1}_S$ need not be Riemann integrable, so we must keep track of how much of S lies in each small cube.

The closed cubes Q_i may overlap on their faces. Assign each shared face to one of the cubes that contains it. This gives disjoint Borel sets $P_i \subseteq Q_i$ whose union is R. Put $S_i = S \cap P_i$. Then S is the disjoint union of the sets S_i, and each S_i lies in a cube where (106) applies.

The map γ is injective: its last k coordinates recover v. Hence the images $\gamma(S_i)$ are also disjoint. They are Borel sets, since γ is a homeomorphism from V onto its graph and that graph is Borel in $\mathbb{R}^{n+k}$. We may therefore add their Hausdorff measures. Summing (106) gives

$$(1-\varepsilon_d)^k \sum_i J_k\gamma(v_i)|S_i| \le \mathcal{H}^k(\gamma(S)) \le (1+\varepsilon_d)^k \sum_i J_k\gamma(v_i)|S_i|.$$

It remains to identify the sum in this inequality. On S_i, we have replaced $J_k\gamma(v)$ by its value at v_i. Every point of S_i is within distance d of v_i. Since $J_k\gamma$ is uniformly continuous on R, the error in this replacement tends uniformly to zero as $d \to 0$. More precisely,

$$\left| \sum_i J_k\gamma(v_i)|S_i| - \int_S J_k\gamma(v)\,dv \right| \le |R| \sup_{\substack{v,w \in R \\ |v-w| \le d}} |J_k\gamma(v) - J_k\gamma(w)| \longrightarrow 0.$$

We also have $\varepsilon_d \to 0$. Letting $d \to 0$ in the preceding upper and lower bounds proves

$$\mathcal{H}^k(\gamma(S)) = \int_S J_k\gamma(v)\,dv$$

for every Borel set $S \subseteq R$.

To pass from one cube to all of V, cover V by countably many closed cubes R_j contained in V. Assign points belonging to more than one cube to the first such cube. This partitions any Borel set $S \subseteq V$ into disjoint Borel pieces, each contained in one R_j. The formula holds on each piece by

what we have just proved. Their graph images are disjoint, so countable additivity of $\mathcal{H}^k$ and of the integral gives

$$\mathcal{H}^k(\gamma(S)) = \int_S J_k\gamma(v)\,dv.$$

This is (101). □

4. Coordinates near a full-rank point

For now suppose that F is C^1. Let x_0 satisfy $\operatorname{rank} DF(x_0) = n$. After permuting coordinates, write $x = (u, v) \in \mathbb{R}^n \times \mathbb{R}^k$ so that, with

$$DF = (A \;\; B), \qquad A = D_uF, B = D_vF$$

the matrix $A(x_0)$ is invertible. Consider

$$\Phi(u, v) = (F(u, v), v), \qquad D\Phi = \begin{pmatrix} A & B \\ 0 & I_k \end{pmatrix}.$$

The inverse function theorem gives a neighbourhood U of x_0 on which Φ is a diffeomorphism onto an open set V. Write its inverse as

$$\Psi(y, v) = (u(y, v), v).$$

For fixed y, the map $v \mapsto \Psi(y, v)$ parametrizes the part of $F^{-1}(y)$ in U.

5. The tangent space and the fibre area factor

The identity $F(u(y, v), v) = y$ holds throughout V. Differentiate it with respect to v, keeping y fixed. Although y is fixed, $u(y, v)$ varies, and the chain rule gives

$$AD_vu + B = 0.$$

Consequently,

$$D_vu = -A^{-1}B, \qquad D_v\Psi = \begin{pmatrix} -A^{-1}B \\ I_k \end{pmatrix}. \tag{107}$$

For a parameter direction $\dot{v}$, the ambient vector $(-A^{-1}B\dot{v}, \dot{v})$ is tangent to the fibre: applying $DF = (A \;\; B)$ to it gives zero.

Observe that $v \mapsto \Psi(y, v)$ is a parametrisation of the fibre $F^{-1}(y)$ lying inside the present chart. By Lemma 10.21, the area factor for the fibre parametrization (with fixed fibre y) is

$$G(y, v) = \sqrt{\det\big((D_v\Psi)^{\mathsf{T}} D_v\Psi\big)} = \sqrt{\det\big(I_k + B^{\mathsf{T}}A^{-\mathsf{T}}A^{-1}B\big)}.$$

The identity $\det(I + ST) = \det(I + TS)$ gives

$$\begin{aligned} G(y, v)^2 &= \det\big(I_n + A^{-1}BB^{\mathsf{T}}A^{-\mathsf{T}}\big) \\ &= \frac{\det(AA^{\mathsf{T}} + BB^{\mathsf{T}})}{(\det A)^2}. \end{aligned}$$

As $DF = (A \;\; B)$, we have shown that

$$G(y, v) = \frac{JF(\Psi(y, v))}{|\det A(\Psi(y, v))|}. \tag{108}$$

Here and below, matrices such as A and B on the right are evaluated at $\Psi(y, v)$.

6. The formula within one chart

Here $U \subseteq \mathbb{R}^N$ is a neighbourhood on which

$$\Phi : U \longrightarrow V, \qquad \Phi(u, v) = (F(u, v), v),$$

is a C^1 diffeomorphism onto an open set $V \subseteq \mathbb{R}^n \times \mathbb{R}^k$. Its inverse is

$$\Psi : V \longrightarrow U, \qquad \Psi(y, v) = (u(y, v), v).$$

Thus the change of variables replaces $x = (u, v) \in U$ by

$$(y, v) = \Phi(x) = (F(u, v), v).$$

The v-coordinate stays the same, while $F(u, v)$ becomes the y-coordinate.

Let $E \subseteq U$ be Borel. In the new coordinates, E corresponds to

$$\Phi(E) = \{(y, v) \in V : \Psi(y, v) \in E\}.$$

Since

$$D\Phi(u, v) = \begin{pmatrix} A(u, v) & B(u, v) \\ 0 & I_k \end{pmatrix}, \qquad A = D_u F,$$

the block-triangular form gives

$$\det D\Phi(u, v) = \det A(u, v).$$

The derivative of Ψ is the inverse of the derivative of Φ at the corresponding point. Hence

$$|\det D\Psi(y, v)| = \frac{1}{|\det A(\Psi(y, v))|}.$$

We now apply the ordinary change-of-variables theorem to the nonnegative function $\mathbf{1}_E JF$:

$$\begin{aligned} \int_E JF(x)\,dx &= \int_U \mathbf{1}_E(x) JF(x)\,dx \\ &= \int_V \mathbf{1}_E(\Psi(y, v)) JF(\Psi(y, v)) |\det D\Psi(y, v)|\,dy\,dv \\ &= \int_V \mathbf{1}_E(\Psi(y, v)) \frac{JF(\Psi(y, v))}{|\det A(\Psi(y, v))|}\,dy\,dv. \end{aligned}$$

By (108), the quotient in the last line is $G(y, v)$. Therefore

$$\int_E JF(x)\,dx = \int_V \mathbf{1}_E(\Psi(y, v)) G(y, v)\,dy\,dv. \tag{109}$$

Since the integrands are nonnegative, the change-of-variables formula also applies when the integrals are infinite. For fixed y, put

$$S_y = \{v : (y, v) \in V,\ \Psi(y, v) \in E\}.$$

Then $\Psi(y, S_y) = E \cap F^{-1}(y)$. Tonelli's theorem and Lemma 10.21 turn (109) into

$$\int_E JF(x)\,dx = \int_{\mathbb{R}^n} \left(\int_{S_y} G(y, v)\,dv \right) dy = \int_{\mathbb{R}^n} \mathcal{H}^k(E \cap F^{-1}(y))\,dy. \tag{110}$$

7. The full-rank set

Let $R = \{x : \operatorname{rank} DF(x) = n\}$. This set is open. The neighbourhoods used in Step 4 cover R, and we may choose a countable subcover $U_1, U_2, \ldots$. For a Borel set $E \subseteq R$, put

$$E_1 = E \cap U_1, \qquad E_j = (E \cap U_j) \setminus \bigcup_{i<j} U_i \quad (j \geq 2).$$

These are disjoint Borel sets, each contained in one chart, and their union is E. Apply (110) to each E_j and sum. For fixed y, every $E_j \cap F^{-1}(y)$ is Borel, so $\mathcal{H}^k$ is countably additive on these disjoint

sets. The fibre integrals are measurable by the chart formula. Countable additivity and monotone convergence therefore give

$$\int_E JF(x)\,dx = \int_{\mathbb{R}^n} \mathcal{H}^k(E \cap F^{-1}(y))\,dy. \tag{111}$$

8. The image of a small cube meeting the critical set

Let $C = \{x : \operatorname{rank} DF(x) < n\}$. On C, one has $JF = 0$. To establish the formula there, we shall prove

$$\mathcal{H}^k(C \cap F^{-1}(y)) = 0 \quad \text{for almost every } y. \tag{112}$$

We do not need $F(C)$ itself to have measure zero.

Fix a compact set $K \subseteq C$ and $0 < \varepsilon \leq 1$. Since K is compact, its closed 1-neighbourhood

$$K_1 := \{x \in \mathbb{R}^N : \operatorname{dist}(x, K) \leq 1\}$$

is compact. The derivative DF is uniformly continuous on K_1. Consequently, there is $\delta > 0$ such that

$$x, z \in K_1, \quad |x - z| < \delta \quad \Longrightarrow \quad \|DF(x) - DF(z)\| < \frac{\varepsilon}{\sqrt{N}}.$$

Let Q be a cube of side length ℓ meeting K, and choose $a \in K \cap Q$. If $b \in K \cap Q$, then

$$|b - a| \leq \sqrt{N}\,\ell.$$

Assume that ℓ is small enough that $\sqrt{N}\,\ell < \min\{1, \delta\}$. For $0 \leq t \leq 1$, the point $a + t(b-a)$ lies in K_1 and satisfies

$$|a + t(b-a) - a| \leq \sqrt{N}\,\ell < \delta.$$

Applying the fundamental theorem of calculus to $t \mapsto F(a + t(b-a))$, we obtain

$$\begin{aligned} F(b) - F(a) &= \int_0^1 DF(a + t(b-a))(b-a)\,dt \\ &= DF(a)(b-a) + \int_0^1 \big[DF(a + t(b-a)) - DF(a)\big](b-a)\,dt. \end{aligned}$$

Define the last integral to be $R_a(b)$. Uniform continuity gives

$$|R_a(b)| \leq \int_0^1 \frac{\varepsilon}{\sqrt{N}} |b-a|\,dt \leq \varepsilon\ell.$$

Thus

$$F(b) = F(a) + DF(a)(b-a) + R_a(b), \qquad |R_a(b)| \leq \varepsilon\ell \quad (b \in K \cap Q). \tag{113}$$

The choice of ℓ depends on ε and K, but not on Q, a, or b. Hence the estimate holds uniformly over all cubes of sufficiently small side length that meet K.

Let P_a be the orthogonal projection of $\mathbb{R}^n$ onto $\operatorname{im} DF(a)$, and put $m = \operatorname{rank} DF(a) < n$. Both a and b lie in $K \cap Q$; we now examine the image point $F(b)$. Since $DF(a)(b-a)$ lies in $\operatorname{im} DF(a)$, (113) gives

$$|(I - P_a)(F(b) - F(a))| = |(I - P_a)R_a(b)| \leq \varepsilon\ell.$$

Thus, perpendicular to $\operatorname{im} DF(a)$, the image $F(K \cap Q)$ has thickness at most $\varepsilon\ell$. In directions parallel to $\operatorname{im} DF(a)$, we have

$$|P_a(F(b) - F(a))| \leq \|DF(a)\|\,|b-a| + |R_a(b)| \leq C\ell,$$

where C is independent of Q and a, since DF is bounded near K.

In orthogonal coordinates corresponding to $\mathbb{R}^n = \operatorname{im} DF(a) \oplus (\operatorname{im} DF(a))^{\perp}$, the set $F(K \cap Q)$ is therefore contained in the product of an m-dimensional ball of radius $C\ell$ and an $(n-m)$-dimensional ball of radius $\varepsilon\ell$. Consequently,

$$|F(K \cap Q)| \le C'\ell^m(\varepsilon\ell)^{n-m} \le C'\varepsilon\ell^n \qquad (0 < \varepsilon \le 1). \tag{114}$$

9. The critical fibres are null almost everywhere

Let $K \subseteq C$ be compact. We shall prove that

$$\mathcal{H}^k(K \cap F^{-1}(y)) = 0 \quad \text{for almost every } y \in \mathbb{R}^n.$$

The idea is to cover K by small cubes. A cube of side length ℓ can contribute at most $C\ell^k$ to a covering of any one fibre. By (114), however, the set of values y whose fibre meets that cube has measure at most $C\varepsilon\ell^n$.

For $\delta > 0$, put $K_y = K \cap F^{-1}(y)$ and define $c_\delta(y)$ to be the infimum of

$$\sum_{i=1}^{s} \sigma_k (\operatorname{diam} U_i)^k$$

over finite families of open sets U_i of diameter less than δ that cover K_y. We use the empty family when K_y is empty. We first check that c_δ is measurable.

Suppose that $c_\delta(y) < t$. Choose an admissible family $U_1, \dots, U_s$ whose sum is less than t. The compact set

$$K \setminus \bigcup_{i=1}^{s} U_i$$

has an image under F that does not contain y. Since that image is compact, it also excludes all y' sufficiently close to y. Thus the same family covers $K_{y'}$ for every such y', and $c_\delta(y') < t$. Hence $\{y : c_\delta(y) < t\}$ is open. In particular, c_δ is upper semicontinuous.

For each fixed y, the quantity $c_\delta(y)$ tends to $\mathcal{H}^k(K_y)$ as $\delta \downarrow 0$. Indeed,

$$\mathcal{H}^k_\delta(K_y) \le c_\delta(y) \le \mathcal{H}^k_{\delta/2}(K_y).$$

The first inequality holds because every finite open cover used to define $c_\delta(y)$ is an admissible Hausdorff cover. For the second, start with any countable cover used to define $\mathcal{H}^k_{\delta/2}(K_y)$. Enlarge its sets slightly to open sets of diameter less than δ, increasing the covering sum by an arbitrarily small amount. These open sets cover the compact set K_y, so finitely many suffice. Taking infima proves the second inequality. It follows that

$$c_\delta(y) \uparrow \mathcal{H}^k(K_y) \quad \text{as } \delta \downarrow 0. \tag{115}$$

We now estimate the integral of c_δ. Fix $\delta > 0$ and $0 < \varepsilon < 1$. Cover K by the closed cubes Q of a grid of side length ℓ that meet K. There are finitely many such cubes, and their number is at most $C_K\ell^{-N}$ when $\ell \le 1$. Choose ℓ sufficiently small that (114) holds for every one of them. We also require each Q to have an open enlargement U_Q with

$$Q \subseteq U_Q, \qquad \operatorname{diam} U_Q < \delta, \qquad \operatorname{diam} U_Q \le C\ell.$$

Fix y. If $x \in K_y$, then x lies in some cube Q in our collection, and $y = F(x) \in F(K \cap Q)$. Consequently, the sets U_Q for which $y \in F(K \cap Q)$ cover K_y. Their covering sum gives

$$c_\delta(y) \le C\ell^k \sum_{Q} \mathbf{1}_{F(K\cap Q)}(y).$$

Each $F(K\cap Q)$ is compact, since both K and Q are closed and bounded. We may therefore integrate the preceding inequality. Using (114) and $k+n=N$, we obtain

$$\begin{aligned}\int_{\mathbb{R}^n} c_\delta(y)\,dy &\le C\ell^k \sum_Q |F(K\cap Q)| \\ &\le C\varepsilon \sum_Q \ell^{k+n} \\ &= C\varepsilon \sum_Q \ell^N \le C_K\varepsilon.\end{aligned}$$

The constant C_K is independent of ε: we are free to decrease ℓ whenever ε decreases. Since $\varepsilon>0$ was arbitrary,

$$\int_{\mathbb{R}^n} c_\delta(y)\,dy = 0 \quad \text{for every } \delta>0.$$

Apply (115) along $\delta=2^{-j}$ and use monotone convergence. This gives

$$\int_{\mathbb{R}^n} \mathcal{H}^k(K\cap F^{-1}(y))\,dy = 0,$$

and proves the assertion for K.

Finally, $C=\{x : \operatorname{rank} DF(x) < n\}$ is closed, because it is the common zero set of the $n\times n$ minors of the continuous matrix-valued function DF. Apply the result just proved to $K_j = C\cap[-j,j]^N$ for each $j\ge 1$. Outside a single null set of values y, every $K_j\cap F^{-1}(y)$ has zero $\mathcal{H}^k$-measure. As the union of the K_j is C, countable subadditivity yields

$$\mathcal{H}^k(C\cap F^{-1}(y)) = 0 \quad \text{for almost every } y.$$

10. Completion of the C^1 case

For any Borel E, split it into $E\cap R$ and $E\cap C$. Equation (111) gives the formula on $E\cap R$. On $E\cap C$, the left-hand side is zero because $JF=0$, and the right-hand side is zero by (112). Thus (99) holds whenever F is C^1.

11. A null set is null on almost every fibre

We return to a Lipschitz map $F:\mathbb{R}^N\to\mathbb{R}^n$, with $k=N-n$.

Lemma 10.22 (Null sets and fibres)**.** *If $Z\subseteq\mathbb{R}^N$ has Lebesgue measure zero, then*

$$\mathcal{H}^k(Z\cap F^{-1}(y)) = 0 \quad \textit{for almost every } y\in\mathbb{R}^n.$$

Proof. We first explain how to cover Z by small balls. Choose a Borel null set E containing Z. By the identification of $\mathcal{H}^N$ with Lebesgue measure on Borel subsets of $\mathbb{R}^N$, we have $\mathcal{H}^N(E)=0$.

Put $\delta_j=2^{-j}$. The definition of $\mathcal{H}^N(E)=0$ gives a countable cover $E\subseteq\bigcup_i U_{i,j}$ such that

$$\operatorname{diam} U_{i,j} < \frac{\delta_j}{4}, \qquad \sum_i (\operatorname{diam} U_{i,j})^N < 2^{-j-N-2}.$$

Discard any empty sets. Choose $a_{i,j}\in U_{i,j}$ and positive numbers $\eta_{i,j}<\delta_j/8$ with

$$\sum_i \eta_{i,j}^N < 2^{-j-N-2}.$$

Set

$$r_{i,j} = \operatorname{diam} U_{i,j} + \eta_{i,j}.$$

Then $U_{i,j} \subseteq B(a_{i,j}, r_{i,j})$ and $2r_{i,j} < \delta_j$. Moreover, using $(s+t)^N \le 2^{N-1}(s^N + t^N)$, we obtain

$$\sum_i r_{i,j}^N < 2^{-j}.$$

Thus, for every j, the balls $B(a_{i,j}, r_{i,j})$ cover Z, have diameter less than δ_j, and satisfy this last estimate.

Write $L_0 = \max\{\mathrm{Lip}(F), 1\}$. Fix $y \in \mathbb{R}^n$. If $x \in Z \cap F^{-1}(y)$ belongs to $B(a_{i,j}, r_{i,j})$, then

$$|y - F(a_{i,j})| = |F(x) - F(a_{i,j})| \le \mathrm{Lip}(F)|x - a_{i,j}| < L_0 r_{i,j}.$$

Consequently, the balls $B(a_{i,j}, r_{i,j})$ for which $y \in B(F(a_{i,j}), L_0 r_{i,j})$ cover $Z \cap F^{-1}(y)$. Each has diameter less than δ_j, so

$$\mathcal{H}^k_{\delta_j}(Z \cap F^{-1}(y)) \le h_j(y),$$

where

$$h_j(y) := \sigma_k 2^k \sum_i r_{i,j}^k \mathbf{1}_{B(F(a_{i,j}), L_0 r_{i,j})}(y).$$

When $k = 0$, each ball in the cover has weight one, as prescribed by the definition of $\mathcal{H}^0_{\delta_j}$.

The functions h_j are measurable and nonnegative. Since a ball of radius $L_0 r_{i,j}$ in $\mathbb{R}^n$ has volume $\omega_n L_0^n r_{i,j}^n$, Tonelli's theorem gives

$$\int_{\mathbb{R}^n} h_j(y)\, dy = \sigma_k 2^k \omega_n L_0^n \sum_i r_{i,j}^{k+n} \le C_{N,n} L_0^n 2^{-j}.$$

Here we used $k + n = N$. Summing over j shows that $\sum_j h_j(y) < \infty$ for almost every y. In particular, $h_j(y) \to 0$ at each such y.

For these values of y,

$$0 \le \mathcal{H}^k_{\delta_j}(Z \cap F^{-1}(y)) \le h_j(y) \longrightarrow 0.$$

As $\delta_j \downarrow 0$, the quantities on the left increase to $\mathcal{H}^k(Z \cap F^{-1}(y))$. Their limit must therefore be zero. □

12. Agreement with a C^1 map on a large compact set

We use the following form of C^1 Lusin approximation.

The version of Whitney's C^1 extension theorem used here says the following. Suppose $K \subset \mathbb{R}^N$ is compact, $f : K \to \mathbb{R}^n$ and $A : K \to \mathcal{L}(\mathbb{R}^N, \mathbb{R}^n)$ are continuous, and

$$\sup_{\substack{x,z \in K \\ 0 < |z-x| < t}} \frac{|f(z) - f(x) - A(x)(z-x)|}{|z-x|} \longrightarrow 0 \quad (t \downarrow 0).$$

Then there is $G \in C^1(\mathbb{R}^N, \mathbb{R}^n)$ with $G = f$ and $DG = A$ on K. The vector-valued statement follows by applying the scalar theorem to each coordinate.

Lemma 10.23 (Lusin–Whitney approximation)**.** *Let $Q \subset \mathbb{R}^N$ be a bounded closed cube. For every $\eta > 0$, there are a compact set $K \subseteq Q$ and $G \in C^1(\mathbb{R}^N, \mathbb{R}^n)$ such that*

$$|Q \setminus K| < \eta, \qquad F = G, \quad DF = DG \quad \text{on } K. \tag{116}$$

Proof. Rademacher's theorem gives DF almost everywhere. Lusin's theorem allows us to restrict to a compact set of almost full measure in Q on which DF is continuous. For points x at which F is differentiable, put

$$e_m(x) = \sup_{0 < |h| < 1/m} \frac{|F(x+h) - F(x) - DF(x)h|}{|h|}.$$

These are measurable functions: the supremum may be taken over a countable dense set of nonzero increments. Differentiability gives $e_m(x) \to 0$ almost everywhere in Q. By Egorov's theorem we

may make this convergence uniform on a measurable set of almost full measure, and then use inner regularity to restrict to a compact subset. Intersect it with the Lusin set and call the resulting compact set K. We may make $|Q \setminus K| < \eta$.

On K, the derivative DF is continuous, and uniform convergence of the e_m implies

$$\sup_{\substack{x,z\in K\\ 0<|z-x|<t}} \frac{|F(z)-F(x)-DF(x)(z-x)|}{|z-x|} \longrightarrow 0 \qquad (t \downarrow 0).$$

These are the C^1 Whitney compatibility conditions for the prescribed values $F|_K$ and derivatives $DF|_K$. The C^1 Whitney extension theorem supplies G satisfying (116). □

Let $E \subseteq K$ be Borel, and let G be as in the lemma. On E, $JF = JG$. Moreover, since $F = G$ at every point of E,

$$E \cap F^{-1}(y) = E \cap G^{-1}(y) \quad \text{for every } y.$$

The C^1 formula for G therefore gives

$$\int_E JF(x)\,dx = \int_{\mathbb{R}^n} \mathcal{H}^k(E \cap F^{-1}(y))\,dy. \tag{117}$$

13. Covering almost all of $\mathbb{R}^N$

Put $Q_m = [-m, m]^N$. For each m, j, apply Lemma 10.23 on Q_m with $\eta = 1/j$, and call its compact agreement set $K_{m,j}$. Let

$$D = \bigcup_{m,j} K_{m,j}.$$

For every fixed m,

$$|Q_m \setminus D| \le |Q_m \setminus K_{m,j}| < \frac{1}{j} \quad \text{for every } j.$$

Thus $Z := \mathbb{R}^N \setminus D$ is a Lebesgue null set.

Enumerate the sets $K_{m,j}$ as $K_1, K_2, \ldots$, and form disjoint Borel pieces

$$D_1 = K_1, \qquad D_i = K_i \setminus \bigcup_{h<i} K_h \quad (i \ge 2).$$

For a Borel set E, apply (117) to $E \cap D_i \subseteq K_i$ and sum over i. This proves (99) on $E \cap D$. The remaining set $E \cap Z$ contributes zero to the left-hand side, and by Lemma 10.22 it has zero $\mathcal{H}^k$-measure on almost every fibre. This proves (99) for every Borel E.

14. Measurable sets and the weighted formula

A Lebesgue measurable set differs from a Borel set by a subset of a Lebesgue null set. Lemma 10.22 therefore extends (99) to Lebesgue measurable E. The fibre intersections are $\mathcal{H}^k$-measurable for almost every y, because they differ there from Borel fibre intersections by $\mathcal{H}^k$-null sets.

For a nonnegative Borel simple function g, apply (99) to each of its characteristic-function terms. If $g_j \uparrow g$ is a sequence of such simple functions, monotone convergence on each fibre, then in y, and on the left-hand side gives (100) for Borel g. Finally, a nonnegative Lebesgue measurable g agrees with a Borel function outside a Lebesgue null set. By Lemma 10.22, this change does not affect the inner integral for almost every y, and the restriction of g to such a fibre is $\mathcal{H}^k$-measurable. Hence (100) holds for such g as well. □

For $n = 1$, $Df(x)$ is the row vector $\nabla f(x)$, and hence $Jf(x) = |\nabla f(x)|$ almost everywhere. In this case (100) reads

$$\int_{\mathbb{R}^N} g(x)|\nabla f(x)|\,dx = \int_{-\infty}^{\infty} \left(\int_{f^{-1}(t)} g(x)\,d\mathcal{H}^{N-1}(x) \right) dt. \tag{118}$$

We now state several applications of the co-area formula, to convince the reader of its broad usefulness.

One use of the co-area formula is to link geometric inequalities to analytic ones. For example, when $n \geq 2$, the sharp isoperimetric inequality (98) can be combined with the co-area formula and the layer-cake representation to give the sharp Sobolev inequality

$$\|\varphi\|_{L^{\frac{n}{n-1}}(\mathbb{R}^n)} \leq \frac{1}{n\omega_n^{1/n}} \int_{\mathbb{R}^n} |\nabla \varphi(x)| \, dx, \tag{119}$$

where $\varphi \in C_c^\infty(\mathbb{R}^n)$. For almost every regular value $t > 0$, the boundary of $\{|\varphi| > t\}$ is the union of the level sets $\varphi^{-1}(t)$ and $\varphi^{-1}(-t)$.

Another immediate consequence of the co-area formula is integration in polar coordinates:

Proposition 10.24 (Integration in polar coordinates)**.** *If $g : \mathbb{R}^n \to \mathbb{R}$ is integrable, then*

$$\int_{\mathbb{R}^n} g(x) \, dx = \int_0^\infty \left(\int_{|x|=r} g \, d\mathcal{H}^{n-1} \right) \, dr. \tag{120}$$

The proof is immediate by setting $f(x) = |x|$ in (118).

Another application is the following weak variant of Sard's theorem[45]:

Proposition 10.25 ("Weak" Sard theorem)**.** *Suppose $n \geq m$, $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1, and $U \subseteq \mathbb{R}^n$ is open. Then for $\mathcal{H}^m$ a.e. $y \in \mathbb{R}^m$, $U \cap f^{-1}(y)$ is the union of a $(n-m)$-dimensional C^1 submanifold of U and a set of $\mathcal{H}^{n-m}$-measure zero. More formally, define*

$$C = \{x \in U : \operatorname{rank}(Df(x)) < m\}.$$

Then $\mathcal{H}^{n-m}\left(f^{-1}(y) \cap C\right) = 0$ for a.e. y and $(U \cap f^{-1}(y)) \setminus C$ is a $(n-m)$-dimensional C^1 submanifold of U.

The proof is left as an exercise. (Hint: For the first claim, observe that $Jf = 0$ iff $\operatorname{rank} Df(x) < m$. The second claim is standard, and follows from the implicit function theorem).

Another application of the co-area formula is the celebrated Faber-Krahn inequality[46].

REFERENCES

[AF03] Robert A. Adams and John J. F. Fournier. *Sobolev spaces*, volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, second edition, 2003.

[Fal86] K. J. Falconer. *The geometry of fractal sets*, volume 85 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1986.

[Fed69] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York, Inc., New York, 1969.

[Fol99] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[Gar73] D. J. H. Garling. A "short" proof of the Riesz representation theorem. *Proc. Cambridge Philos. Soc.*, 73:459–460, 1973.

[Hö90] Lars Hörmander. *The analysis of linear partial differential operators. I*, volume 256 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 1990. Distribution theory and Fourier analysis.

[Har83] Donald G. Hartig. The Riesz representation theorem revisited. *Amer. Math. Monthly*, 90(4):277–280, 1983.

[LL01] Elliott H. Lieb and Michael Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2001.

[45] The full version will be a topic of presentation. The version given here has a weaker conclusion, but observe the fact that we also have a weaker hypothesis!

[46] For more details, see https://cuhkmath.wordpress.com/2017/09/09/faber-krahn-inequality/.

[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[Roy88] H. L. Royden. *Real analysis*. Macmillan Publishing Company, New York, third edition, 1988.

[Rud87] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[Sim83] Leon Simon. *Lectures on geometric measure theory*, volume 3 of *Proceedings of the Centre for Mathematical Analysis, Australian National University*. Australian National University, Centre for Mathematical Analysis, Canberra, 1983.

[Tay06] Michael E. Taylor. *Measure theory and integration*, volume 76 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2006.