Lecture Notes in Advanced Real Analysis

Eric T. Sawyer

McMaster University, Hamilton, Ontario
E-mail address: sawyer@mcmaster.ca

Abstract. Beginning with Lebesgue integration on the real line, and continuing with Euclidean spaces, the Banach-Tarski paradox, and the Riesz representation theorem on locally compact Hausdorff spaces, these lecture notes examine theories of integration with applications to analysis and differential equations.

Contents

Preface

These notes grew out of lectures given twice a week in a first year graduate course in advanced real analysis at McMaster University September to December 2010. Part 1 consists of a brief review of compactness and continuity. The topics in Part 2 include Lebesgue integration on Euclidean spaces, the Banach-Tarski paradox, the Riesz representation theorem on locally compact Hausdorff spaces, Lebesgue spaces $L^p(\mu)$, Banach and Hilbert spaces, complex measures and the Radon-Nikodym theorem, and Fubini's theorem. Applications to differential equations will be forthcoming in Part 3. Sources include books by Rudin [**3**], [**4**] and [**5**], and books by Stein and Shakarchi [**6**] and [**7**]. Special topics are covered in Bartle and Sherbert [**1**] and Wagon [**8**].

Part 1

Topology of Euclidean spaces

We begin Part 1 by reviewing some of the theory of compact sets and continuity of functions in Euclidean spaces $\mathbb{R}^n$. We assume the reader is already familiar with the notions of sequence, open, closed, countable and uncountable, and is comfortable with elementary properties of limits, continuity and differentiability of functions.

CHAPTER 1

Compact sets

Let $\mathbb{R}$ be the set of real numbers equipped with the usual field and order operations, and the least upper bound property. Denote by $\mathbb{R}^n$ the n-dimensional Euclidean space

$$\mathbb{R}^n = \{(x_1, x_2, ..., x_n) \in \mathbb{R} \times \mathbb{R} \times ... \times \mathbb{R} \ (n \text{ times})\}$$

equipped with the usual vector addition and scalar and inner products

$$\begin{aligned} x + y &\equiv (x_1 + y_1, x_2 + y_2, ..., x_n + y_n), \\ \lambda x &= (\lambda x_1, \lambda x_2, ..., \lambda x_n) \\ x \cdot y &\equiv x_1 y_1 + x_2 y_2 + ... + x_n y_n, \end{aligned}$$

if $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n)$ and $\lambda \in \mathbb{R}$.

We begin with the single most important property that a subset of Euclidean space can have, namely *compactness*. In a sense, compact subsets share the most important topological properties enjoyed by *finite* sets. It turns out that the most basic of these properties is rather abstract looking at first sight, but arises so often in applications and subsequent theory that we will use it as the definition of compactness. But first we introduce some needed terminology.

Let E be a subset of $\mathbb{R}^n$. A collection $\mathcal{G} \equiv \{G_\alpha\}_{\alpha \in A}$ of subsets G_α of $\mathbb{R}^n$ is said to be an *open cover* of E if

$$\text{each } G_\alpha \text{ is open and } E \subset \bigcup_{\alpha \in A} G_\alpha.$$

A *finite* subcover (relative to the open cover $\mathcal{G}$ of E) is a finite collection $\{G_{\alpha_k}\}_{k=1}^n$ of the open sets G_α that still covers E:

$$E \subset \bigcup_{k=1}^n G_{\alpha_k}.$$

For example, the collection $\mathcal{G} = \left\{\left(\frac{1}{n}, 1 + \frac{1}{n}\right)\right\}_{n=1}^\infty$ of open intervals in $\mathbb{R}$ form an open cover of the interval $E = \left(\frac{1}{8}, 2\right)$, and $\left\{\left(\frac{1}{n}, 1 + \frac{1}{n}\right)\right\}_{n=1}^8$ is a finite subcover. Draw a picture! However, $\mathcal{G}$ is also an open cover of the interval $E = (0, 2)$ for which there is *no* finite subcover since $\frac{1}{m} \notin \left(\frac{1}{n}, 1 + \frac{1}{n}\right)$ for all $1 \le n \le m$.

Definition 1. *A subset E of $\mathbb{R}^n$ is* compact *if every open cover of E has a finite subcover.*

Example 1. *Clearly every finite set is compact. On the other hand, the interval $(0, 2)$ is not compact since $\mathcal{G} = \left\{\left(\frac{1}{n}, 1 + \frac{1}{n}\right)\right\}_{n=1}^\infty$ is an open cover of $(0, 2)$ that does* not *have a finite subcover.*

The above example makes it clear that all we need is one 'bad' cover as witness to the failure of a set to be compact. On the other hand, in order to show that an infinite set *is* compact, we must often work much harder, namely we must show that given *any* open cover, there is *always* a finite subcover. It will obviously be of great advantage if we can find simpler criteria for a set to be compact, and this will be carried out below. For now we will content ourselves with giving one simple example of an *infinite* compact subset of the real numbers (even of the rational numbers).

Example 2. *The set* $K \equiv \{0\} \cup \left\{\frac{1}{k}\right\}_{k=1}^{\infty}$ *is compact in* $\mathbb{R}$. *Indeed, suppose that* $\mathcal{G} \equiv \{G_\alpha\}_{\alpha \in A}$ *is an open cover of* K. *Then at least one of the open sets in* $\mathcal{G}$ *contains* 0, *say* G_{α_0}. *Since* G_{α_0} *is open, there is* $r > 0$ *such that*

$$B(0,r) \subset G_{\alpha_0}.$$

Now comes the crux of the argument: there are only finitely *many points* $\frac{1}{k}$ *that lie outside* $B(0,r)$, *i.e.* $\frac{1}{k} \notin B(0,r)$ *if and only if* $k \leq \left[\frac{1}{r}\right] \equiv n$. *Now choose* G_{α_k} *to contain* $\frac{1}{k}$ *for each* k *between* 1 *and* n *inclusive (with possible repetitions). Then the finite collection of open sets* $\{G_{\alpha_0}, G_{\alpha_1}, G_{\alpha_2}, ..., G_{\alpha_n}\}$ *(after removing repetitions) constitute a finite subcover relative to the open cover* $\mathcal{G}$ *of* K. *Thus we have shown that* every *open cover of* K *has a finite subcover.*

It is instructive to observe that $K = \overline{E}$ where $E = \left\{\frac{1}{k}\right\}_{k=1}^{\infty}$ is *not* compact (since the pairwise disjoint balls $B\left(\frac{1}{k}, \frac{1}{4k^2}\right) = \left(\frac{1}{k} - \frac{1}{4k^2}, \frac{1}{k} + \frac{1}{4k^2}\right)$ cover E one point at a time). Thus the addition of the single limit point 0 to the set E resulted in making the union compact. The argument given as proof in the above example serves to illustrate the sense in which the set K is topologically 'almost' a finite set.

As a final example to illustrate the concept of compactness, we show that any *unbounded* set in $\mathbb{R}^n$ fails to be compact. We say that a subset E of $\mathbb{R}^n$ is *bounded* if there is some ball $B(x,r)$ in $\mathbb{R}^n$ that contains E. So now suppose that E is unbounded. Fix a point $x \in \mathbb{R}^n$ and consider the open cover $\{B(x,n)\}_{n=1}^{\infty}$ of E (this is actually an open cover of $\mathbb{R}^n$). Now if there were a finite subcover, say $\{B(x,n_k)\}_{k=1}^{N}$ where $n_1 < n_2 < ... < n_N$, then because the balls are increasing,

$$E \subset \bigcup_{k=1}^{N} B(x,n_k) = B(x,n_N),$$

which contradicts the assumption that E is unbounded. We record this fact in the following lemma.

Lemma 1. *A compact subset of* $\mathbb{R}^n$ *is bounded.*

Remark 1. *We can now preview one of the major themes in our development of analysis. The Least Upper Bound Property of the real numbers will lead directly to the following beautiful characterization of compactness in the metric space* $\mathbb{R}$ *of real numbers, the* Heine-Borel theorem: *a subset* K *of* $\mathbb{R}$ *is compact if and only if* K *is closed and bounded.*

1. Properties of compact sets

We now prove a number of properties that hold for compact sets in Euclidean space $\mathbb{R}^n$.

Lemma 2. *If* K *is a compact subset of* $\mathbb{R}^n$, *then* K *is a closed subset of* $\mathbb{R}^n$.

Proof: We show that K^c is open. So fix a point $x \in K^c$. For each point $y \in K$, consider the ball $B(y, r_y)$ with

$$r_y \equiv \frac{1}{2} d(x, y). \tag{1.1}$$

Since $\{B(y, r_y)\}_{y \in K}$ is an open cover of the compact set K, there is a *finite* subcover $\{B(y_k, r_{y_k})\}_{k=1}^n$ with of course $y_k \in K$ for $1 \leq k \leq n$. Now by the triangle inequality and (1.1) it follows that

$$B(x, r_{y_k}) \cap B(y_k, r_{y_k}) = \emptyset, \qquad 1 \leq k \leq n. \tag{1.2}$$

Indeed, if the intersection on the left side of (1.2) contained a point z then we would have the contradiction

$$d(x, y_k) \leq d(x, z) + d(z, y_k) < r_{y_k} + r_{y_k} = d(x, y_k).$$

Now we simply take $r = \min \{r_{y_k}\}_{k=1}^n > 0$ and note that $B(x, r) \subset B(x, r_{y_k})$ so that

$$\begin{aligned} B(x, r) \cap K &\subset B(x, r) \cap \left(\bigcup_{k=1}^n B(y_k, r_{y_k}) \right) \\ &= \bigcup_{k=1}^n \{B(x, r) \cap B(y_k, r_{y_k})\} \\ &\subset \bigcup_{k=1}^n \{B(x, r_{y_k}) \cap B(y_k, r_{y_k})\} = \bigcup_{k=1}^n \emptyset = \emptyset, \end{aligned}$$

by (1.2). This shows that $B(x, r) \subset K^c$ and completes the proof that K^c is open. Draw a picture of this proof!

LEMMA 3. *If $F \subset K \subset X$ where F is closed in $\mathbb{R}^n$ and K is compact, then F is compact.*

Proof: Let $\mathcal{G} = \{G_\alpha\}_{\alpha \in A}$ be an open cover of F. We must construct a finite subcover $\mathcal{S}$ of F. Now $\mathcal{G}^* = \{F^c\} \cup \mathcal{G}$ is an open cover of K. By compactness of K there is a finite subcover S^* of $\mathcal{G}^*$ that consists of sets from $\mathcal{G}$ and possibly the set F^c. However, if we drop the set F^c from the subcover S^* the resulting finite collection of sets $\mathcal{S}$ from $\mathcal{G}$ is still a cover of F (although not neccessarily of K), and provides the required finite subcover of F.

COROLLARY 1. *If F is closed and K is compact, then $F \cap K$ is compact.*

Proof: We have that K is closed by Lemma 2, and so $F \cap K$ is closed. Now $F \cap K \subset K$ and so Lemma 3 shows that $F \cap K$ is compact.

REMARK 2. *With respect to unions, compact sets behave like finite sets, namely the union of finitely many compact sets is compact. Indeed, suppose K and L are compact subsets of a metric space, and let $\{G_\alpha\}_{\alpha \in A}$ be an open cover of $K \cup L$. Then there is a finite subcover $\{G_\alpha\}_{\alpha \in I}$ of K and also a (usually different) finite subcover $\{G_\alpha\}_{\alpha \in J}$ of L (here I and J are finite subsets of A). But then the union of these covers $\{G_\alpha\}_{\alpha \in I \cup J} = \{G_\alpha\}_{\alpha \in I} \cup \{G_\alpha\}_{\alpha \in J}$ is a finite subcover of $K \cup L$, which shows that $K \cup L$ is compact.*

Now we come to one of the most useful consequences of compactness in applications. A family of sets $\{E_\alpha\}_{\alpha\in A}$ is said to have the *finite intersection property* if

$$\bigcap_{\alpha\in F} E_\alpha \neq \emptyset$$

for every finite subset F of the index set A. For example the family of open intervals $\left\{\left(0,\frac{1}{n}\right)\right\}_{n=1}^{\infty}$ has the finite intersection property despite the fact that the sets have no element in common: $\bigcap_{n=1}^{\infty}\left(0,\frac{1}{n}\right)=\emptyset$. The useful consequence of compactness referrred to above is that this *cannot* happen for compact subsets!

THEOREM 1. *Suppose that* $\{K_\alpha\}_{\alpha\in A}$ *is a family of* compact *sets with the finite intersection property. Then*

$$\bigcap_{\alpha\in A} K_\alpha \neq \emptyset.$$

Proof: Fix a member K_{α_0} of the family $\{K_\alpha\}_{\alpha\in A}$. Assume in order to derive a contradiction that *no* point of K_{α_0} belongs to every K_α. Then the open sets $\{K_\alpha^c\}_{\alpha\in A\setminus\{\alpha_0\}}$ form an open cover of K_{α_0}. By compactness, there is a finite subcover $\{K_\alpha^c\}_{\alpha\in F\setminus\{\alpha_0\}}$ with F finite, so that

$$K_{\alpha_0}\subset\bigcup_{\alpha\in F\setminus\{\alpha_0\}} K_\alpha^c,$$

i.e.

$$K_{\alpha_0}\cap\bigcap_{\alpha\in F\setminus\{\alpha_0\}} K_\alpha=\emptyset,$$

which contradicts our assumption that the finite intersection property holds.

COROLLARY 2. *If* $\{K_n\}_{n=1}^{\infty}$ *is a nonincreasing sequence of nonempty compact sets. i.e.* $K_{n+1}\subset K_n$ *for all* $n\geq 1$*, then*

$$\bigcap_{n=1}^{\infty} K_n \neq \emptyset.$$

THEOREM 2. *If* E *is an infinite subset of a compact set* K*, then* E *has a limit point in* K*.*

Proof: Suppose, in order to derive a contradiction, that *no* point of K is a limit point of E. Then for each $z\in K$, there is a ball $B(z,r_z)$ that contains *at most* one point of E (namely z if z is in E). Thus it is not possible for a finite number of these balls $B(z,r_z)$ to cover the infinite set E. Thus $\{B(z,r_z)\}_{z\in K}$ is an open cover of K that has no finite subcover (since a finite subcover cannot cover even the subset E of K). This contradicts the assumption that K is compact.

The Least Upper Bound Property of the real numbers plays a crucial role in the proof that closed bounded intervals are compact.

THEOREM 3. *The closed interval* $[a,b]$ *is compact (with the usual metric) for all* $a<b$*.*

We give two proofs of this basic theorem. The second proof will be generalized to prove that closed bounded rectangles in $\mathbb{R}^n$ are compact.

Proof #1: Assume for convenience that the interval is the closed unit interval $[0,1]$, and suppose that $\{G_\alpha\}_{\alpha\in A}$ is an open cover of $[0,1]$. Now $1 \in G_\beta$ for some $\beta \in A$ and thus there is $r > 0$ such that $(1-r, 1+r) \subset G_\beta$. With $a = 1 + \frac{r}{2} > 1$ it follows that $\{G_\alpha\}_{\alpha\in A}$ is an open cover of $[0,a]$. Now define

$$E = \{x \in [0,a] : \text{the interval } [0,x] \text{ has a finite subcover}\}.$$

We have E is nonempty ($0 \in E$) and bounded above (by a). Thus $\lambda \equiv \sup E$ exists. We claim that $\lambda > 1$. Suppose for the moment that this has been proved. Then 1 cannot be an upper bound of E and so there is some $\sigma \in E$ satisfying

$$1 < \sigma \leq \lambda.$$

Thus by the definition of the set E it follows that $[0,\sigma]$ has a finite subcover, and hence so does $[0,1]$, which completes the proof of the theorem.

Now suppose, in order to derive a contradiction, that $\lambda \leq 1$. Then there is some open set G_γ with $\gamma \in A$ and also some $s > 0$ such that

$$(\lambda - s, \lambda + s) \subset G_\gamma.$$

Now by the definition of least upper bound, there is some $x \in E$ satisfying $\lambda - s < x \leq \lambda$, and by taking s less than $a - 1$ we can also arrange to have

$$\lambda + s \leq 1 + s < a.$$

Thus there is a finite subcover $\{G_{\alpha_k}\}_{k=1}^n$ of $[0,x]$, and if we include the set G_γ with this subcover we get a finite subcover of $\left[0, \lambda + \frac{s}{2}\right]$. This shows that $\lambda + \frac{s}{2} \in E$, which contradicts our assumption that λ is an upper bound of E, and completes the proof of the theorem.

Proof #2: Suppose, in order to derive a contradiction, that there is an open cover $\{G_\alpha\}_{\alpha\in A}$ of $[a,b]$ that has *no* finite subcover. Then at least one of the two intervals $\left[a, \frac{a+b}{2}\right]$ and $\left[\frac{a+b}{2}, b\right]$ fails to have a finite subcover. Label it $[a_1, b_1]$ so that

$$\begin{aligned} a &\leq a_1 < b_1 \leq b, \\ b_1 - a_1 &= \frac{1}{2}\delta, \end{aligned}$$

where $\delta = b - a$. Next we note that at least one of the two intervals $\left[a_1, \frac{a_1+b_1}{2}\right]$ and $\left[\frac{a_1+b_1}{2}, b_1\right]$ fails to have a finite subcover. Label it $[a_2, b_2]$ so that

$$\begin{aligned} a \leq a_1 &\leq a_2 < b_2 \leq b_1 \leq b, \\ b_2 - a_2 &= \frac{1}{4}\delta. \end{aligned}$$

Continuing in this way we obtain for each $n \geq 2$ an interval $[a_n, b_n]$ such that

$$\begin{aligned} a \leq a_1 \leq ...a_{n-1} &\leq a_n < b_n \leq b_{n-1} ... \leq b_1 \leq b, \\ b_n - a_n &= \frac{1}{2^n}\delta. \end{aligned} \tag{1.3}$$

Now let $E = \{a_n : n \geq 1\}$ and set $x \equiv \sup E$. From (1.3) we obtain that each b_n is an upper bound for E, hence $x \leq b_n$ and we have

$$a \leq a_n \leq x \leq b_n \leq b, \text{ for all } n \geq 1,$$

i.e. $x \in [a_n, b_n]$ for all $n \geq 1$. Now $x \in [a, b]$ and so there is $\beta \in A$ and $r > 0$ such that

$$(x - r, x + r) \subset G_\beta.$$

By the Archimedean property of $\mathbb{R}$ we can choose $n \in \mathbb{N}$ so large that $\frac{1}{r} < n < 2^n$ (it is easy to prove $n < 2^n$ for all $n \in \mathbb{N}$ by induction), and hence

$$[a_n, b_n] \subset (x - r, x + r) \subset G_\beta.$$

But this contradicts our construction that $[a_n, b_n]$ has no finite subcover, and completes the proof of the theorem.

Corollary 3. *A subset K of the real numbers $\mathbb{R}$ is compact* if and only if K *is closed and bounded.*

Proof: Suppose that K is compact. Then K is bounded by Lemma 1 and is closed by Lemma 2. Conversely if K is bounded, then $K \subset [-a, a]$ for some $a > 0$. Now $[-a, a]$ is compact by Theorem 3, and if K is closed, then Lemma 3 shows that K is compact.

Proof #2 of Theorem 3 is easily adapted to prove that closed rectangles

$$R = \prod_{k=1}^{n} [a_k, b_k] = [a_1, b_1] \times \ldots \times [a_n, b_n]$$

in $\mathbb{R}^n$ are compact.

Theorem 4. *The closed rectangle $R = \prod_{k=1}^{n} [a_k, b_k]$ is compact (with the usual metric) for all $a_k < b_k$, $1 \leq k \leq n$.*

Proof: Here is a brief sketch of the proof. Suppose, in order to derive a contradiction, that there is an open cover $\{G_\alpha\}_{\alpha \in A}$ of R that has *no* finite subcover. It is convenient to write R as a product of closed intervals with superscripts instead of subscripts: $R = \prod_{k=1}^{n} \left[a^k, b^k\right]$. Now divide R into 2^n congruent closed rectangles. At least one of them fails to have a finite subcover. Label it $R_1 \equiv \prod_{k=1}^{n} \left[a_1^k, b_1^k\right]$, and repeat the process to obtain a sequence of decreasing rectangles $R_m \equiv \prod_{k=1}^{n} \left[a_m^k, b_m^k\right]$ with

$$\begin{aligned} a^k &\leq a_1^k \leq \ldots a_{m-1}^k \leq a_m^k < b_m^k \leq b_{m-1}^k \ldots \leq b_1^k \leq b^k, \\ b_m^k - a_m^k &= \frac{1}{2^m} \delta^k, \end{aligned}$$

where $\delta^k = b^k - a^k$, $1 \leq k \leq n$. Then if we set $x^k = \sup\left\{a_m^k : m \geq 1\right\}$ we obtain that $x = \left(x^1, \ldots, x^n\right) \in R_m \subset R$ for all m. Thus there is $\beta \in A$, $r > 0$ and $m \geq 1$ such that

$$R_m \subset B(x, r) \subset G_\beta,$$

contradicting our construction that R_m has no finite subcover.

Theorem 5. *Let K be a subset of Euclidean space $\mathbb{R}^n$. Then the following three conditions are equivalent:*

(1) *K is closed and bounded;*
(2) *K is compact;*
(3) *every infinite subset of K has a limit point in K.*

Proof: We prove that (1) implies (2) implies (3) implies (1). If K is closed and bounded, then it is contained in a closed rectangle R, and is thus compact by Theorem 4 and Lemma 3. If K is compact, then every infinite subset of K has a limit point in K by Theorem 2. Finally suppose that every infinite subset of K has a limit point in K. Suppose first, in order to derive a contradiction, that K is not bounded. Then there is a sequence $\{x_k\}_{k=1}^{\infty}$ of points in K with $|x_k| \geq k$ for all k. Clearly the set of points in $\{x_k\}_{k=1}^{\infty}$ is an infinite subset E of K but has no limit point in $\mathbb{R}^n$, hence not in K either. Suppose next, in order to derive a contradiction, that K is not closed. Then there is a limit point x of K that is not in K. Thus each deleted ball $B'\left(x, \frac{1}{k}\right)$ contains some point x_k from K. Again it is clear that the set of points in the sequence $\{x_k\}_{k=1}^{\infty}$ is an infinite subset of K but contains no limit point in K since its only limit point is x and this is not in K.

COROLLARY 4. *Every bounded infinite subset of* $\mathbb{R}^n$ *has a limit point in* $\mathbb{R}^n$.

2. The Cantor set

We now construct the Cantor middle thirds set (1883). This famous fractal set arises as a counterexample to many conjectures in analysis. We start with the closed unit interval $I = I^0 = [0,1]$. Now remove the open middle third $\left(\frac{1}{3}, \frac{2}{3}\right)$ of length $\frac{1}{3}$ and denote the two remaining closed intervals of length $\frac{1}{3}$ by $I_1^1 = \left[0, \frac{1}{3}\right]$ and $I_2^1 = \left[\frac{2}{3}, 1\right]$. Then remove the open middle third $\left(\frac{1}{9}, \frac{2}{9}\right)$ of length $\frac{1}{3^2}$ from $I_1^1 = \left[0, \frac{1}{3}\right]$ and denote the two remaining closed intervals of length $\frac{1}{3^2}$ by I_1^2 and I_2^2. Do the same for I_2^1 and denote the two remaining closed intervals by I_3^2 and I_4^2.

Continuing in this way, we obtain at the k^{th} generation, a collection $\left\{I_j^k\right\}_{j=1}^{2^k}$ of 2^k pairwise disjoint closed intervals of length $\frac{1}{3^k}$. Let $K_k = \bigcup_{j=1}^{2^k} I_j^k$ and set

$$E = \bigcap_{k=1}^{\infty} K_k = \bigcap_{k=1}^{\infty} \left(\bigcup_{j=1}^{2^k} I_j^k \right).$$

Now each set K_k is closed, and hence so is the intersection E. Then E is *compact* by Corollary 3. It also follows from Corollary 2 that E is *nonempty*. Next we observe that by its very construction, E is a fractal satisfying the replication identity

$$3E = E \cup (E+2) = E_1 \cup E_2.$$

Thus the fractal dimension α of the Cantor set E is $\frac{\ln 2}{\ln 3}$. Moreover, E has the property of being *perfect*.

DEFINITION 2. *A subset* E *of* $\mathbb{R}^n$ *is* perfect *if* E *is closed and every point in* E *is a limit point of* E.

To see that the Cantor set is perfect, pick $x \in E$. For each $k \geq 1$ the point x lies in exactly one of the closed intervals I_j^k for some j between 1 and 2^k. Since the length of I_j^k is positive, it is possible to choose a point $x_k \in E \cap I_j^k \setminus \{x\}$. Now the set of points in the sequence $\{x_k\}_{k=1}^{\infty}$ is an infinite subset of E and clearly has x as a limit point. This completes the proof that the Cantor set E is perfect.

By summing the lengths of the removed open middle thirds, we obtain

$$\text{'length'}\left([0,1] \setminus E\right) = \frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + ... = 1,$$

and it follows that E is nonempty, compact and has 'length' $1-1=0$. Another way to exhibit the same phenomenon is to note that for each $k \geq 1$ the Cantor set E is a subset of the closed set K_k which is a union of 2^k intervals each having length $\frac{1}{3^k}$. Thus the 'length' of K_k is $2^k \frac{1}{3^k} = \left(\frac{2}{3}\right)^k$, and the 'length' of E is at most

$$\inf\left\{\left(\frac{2}{3}\right)^k : k \geq 1\right\} = 0.$$

In contrast to this phenomenon that the 'length' of E is quite small, the cardinality of E is quite large, namely E is uncountable, as is every nonempty perfect subset of a metric space with the Heine-Borel property: every closed and bounded subset is compact.

THEOREM 6. *Every nonempty perfect subset of* $\mathbb{R}^n$ *is uncountable.*

Proof: Suppose that P is a nonempty perfect subset of $\mathbb{R}^n$. Since P has a limit point it must be infinite. Now assume, in order to derive a contradiction, that P is countable, say $P = \{x_n\}_{n=1}^{\infty}$. Start with any point $y_1 \in P$ that is not x_1 and the ball $B_1 \equiv B(y_1, r_1)$ where $r_1 = \frac{d(x_1,y_1)}{2}$. We have

$$B_1 \cap P \neq \emptyset \text{ and } x_1 \notin \overline{B_1}.$$

Then there is a point $y_2 \in B_1' \cap P$ that is not x_2 and so we can choose a ball B_2 such that

$$B_2 \cap P \neq \emptyset \text{ and } x_2 \notin \overline{B_2} \text{ and } \overline{B_2} \subset B_1.$$

Indeed, we can take $B_2 = B(y_2, r_2)$ where $r_2 = \frac{\min\{d(x_2,y_2), r_1 - d(y_1,y_2)\}}{2}$. Continuing in this way we obtain balls B_k satisfying

$$B_k \cap P \neq \emptyset \text{ and } x_k \notin \overline{B_k} \text{ and } \overline{B_k} \subset B_{k-1}, \qquad k \geq 2.$$

Now each closed set $\overline{B_k} \cap P$ is nonempty and compact, and so by Corollary 2 we have

$$\bigcap_{k=1}^{\infty} \left(\overline{B_k} \cap P\right) \neq \emptyset, \qquad \text{say } x \in \left(\bigcap_{k=1}^{\infty} \overline{B_k}\right) \cap P.$$

However, by construction we have $x_n \notin \overline{B_n}$ for all n and since the sets $\overline{B_n}$ are decreasing, we see that $x_n \notin \bigcap_{k=1}^{\infty} \overline{B_k}$ for all n; hence $x \neq x_n$ for all n. This contradicts $P = \{x_n\}_{n=1}^{\infty}$ and completes the proof of the theorem.

CHAPTER 2

Continuous functions

We initially examine the connection between continuity and sequences, and after that between continuity and open sets. Central to all of this is the concept of *limit* of a function.

1. Limits

Definition 3. *Suppose that $f : X \to \mathbb{R}^m$ is a function from a subset X of $\mathbb{R}^n$ into $\mathbb{R}^m$. Let $p \in \mathbb{R}^n$ be a limit point of X and suppose that $q \in \mathbb{R}^m$. Then*

$$\lim_{x \to p} f(x) = q$$

if for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$d_{\mathbb{R}^m}(f(x), q) < \varepsilon \text{ whenever } x \in X \setminus \{p\} \text{ and } d_{\mathbb{R}^n}(x, p) < \delta. \tag{1.1}$$

Note that the concept of a limit of f at a point p is only defined when p is a *limit point* of the set X on which f is defined. Do not confuse this notion with the definition of limit of a sequence $s = \{s_n\}_{n=1}^{\infty}$ in $\mathbb{R}^n$. In this latter definition, s is a function from the natural numbers $\mathbb{N}$ into $\mathbb{R}^n$, but the limit point p is replaced by the symbol ∞. Here is a characterization of limit of a function in terms of limits of sequences.

Theorem 7. *Suppose that $f : X \to \mathbb{R}^m$ is a function from a subset X of $\mathbb{R}^n$ into $\mathbb{R}^m$. Let $p \in \mathbb{R}^n$ be a limit point of X and suppose that $q \in \mathbb{R}^m$. Then* $\lim_{x \to p} f(x) = q$ if and only if

$$\lim_{k \to \infty} f(s_k) = q$$

for all sequences $\{s_k\}_{k=1}^{\infty}$ in $X \setminus \{p\}$ such that

$$\lim_{k \to \infty} s_k = p.$$

Proof: Suppose first that $\lim_{x \to p} f(x) = q$. Now assume that $\{s_k\}_{k=1}^{\infty}$ is a sequence in $X \setminus \{p\}$ such that $\lim_{k \to \infty} s_k = p$. Then given $\varepsilon > 0$ there is $\delta > 0$ such that (1.1) holds. Furthermore we can find N so large that $d_{\mathbb{R}^n}(s_k, p) < \delta$ whenever $k \geq N$. Combining inequalities with the fact that $s_k \in E$ gives

$$d_{\mathbb{R}^m}(f(s_k), q) < \varepsilon \text{ whenever } k \geq N,$$

which proves $\lim_{k \to \infty} f(s_k) = q$.

Suppose next that $\lim_{x \to p} f(x) = q$ fails. The negation of Definition 3 is that **there exists** an $\varepsilon > 0$ such that **for every** $\delta > 0$ we have

$$d_{\mathbb{R}^m}(f(x), q) \geq \varepsilon \textbf{ for some } x \in X \setminus \{p\} \text{ with } d_{\mathbb{R}^n}(x, p) < \delta. \tag{1.2}$$

So fix such an $\varepsilon > 0$ and for each $\delta = \frac{1}{k} > 0$ choose a point $s_k \in X \setminus \{p\}$ with $d_{\mathbb{R}^n}(s_k, p) < \frac{1}{k}$. Then $\{s_k\}_{k=1}^{\infty}$ is a sequence in $X \setminus \{p\}$ such that the sequence $\{f(s_k)\}_{k=1}^{\infty}$ does not converge to q - indeed, $d_{\mathbb{R}^m}(f(s_k), q) \geq \varepsilon > 0$ for all $k \geq 1$.

2. Continuity

DEFINITION 4. *Let X be a subset of $\mathbb{R}^n$ and suppose that $f : X \to \mathbb{R}^m$ is a function from X to $\mathbb{R}^m$. Let $p \in X$. Then f is continuous at p if for every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$d_{\mathbb{R}^m}(f(x), f(p)) < \varepsilon \text{ whenever } x \in X \text{ and } d_{\mathbb{R}^n}(x, p) < \delta. \tag{2.1}$$

Note that (2.1) says

$$f(B(p, \delta) \cap X) \subset B(f(p), \varepsilon). \tag{2.2}$$

There are only two possibilities for the point $p \in X$; *either* p is a limit point of X *or* p is isolated in X (a point x in X is isolated in X if there is a deleted ball $B'(x, r)$ that has empty intersection with X). In the case that p is a limit point of X, then f is continuous at p if and only if $\lim_{x \to p} f(x)$ exists and the limit is $f(p)$, i.e.

$$\lim_{x \to p} f(x) = f(p). \tag{2.3}$$

On the other hand, if p is an isolated point of X, then f is automatically continuous at p since (2.1) holds for all $\varepsilon > 0$ with $\delta = r$ where $B'(x, r) \cap X = \emptyset$. From these remarks together with Theorem 7, we immediately obtain the following characterization of continuity in terms of sequences.

THEOREM 8. *Suppose that $f : X \to \mathbb{R}^m$ is a function from a subset X of $\mathbb{R}^n$ into $\mathbb{R}^m$. Let $p \in X$. Then f is continuous at p* if and only if

$$\lim_{k \to \infty} f(s_k) = f(p)$$

for all sequences $\{s_k\}_{k=1}^{\infty}$ in $X \setminus \{p\}$ such that

$$\lim_{k \to \infty} s_k = p.$$

REMARK 3. *The theorem remains true if we permit the sequences $\{s_k\}_{k=1}^{\infty}$ to lie in X rather than in $X \setminus \{p\}$.*

There is an alternate characterization of continuity of $f : X \to Y$ in terms of *relative open* sets, which is particularly useful in connection with compact sets and continuity of inverse functions. Recall that a subset Q of X is said to be *relatively open* in X if there is an open set G in $\mathbb{R}^n$ such that $Q = G \cap X$. We will often drop the adverb *relatively.*

THEOREM 9. *Suppose that $f : X \to Y$ is a function from a subset X of $\mathbb{R}^n$ into a subset Y of $\mathbb{R}^m$. Then f is continuous on X if and only if*

$$f^{-1}(G) \text{ is open in } X \text{ for every } G \text{ that is open in } Y. \tag{2.4}$$

COROLLARY 5. *Suppose that $f : X \to Y$ is a continuous function from a* compact *subset X of $\mathbb{R}^n$ into a subset Y of $\mathbb{R}^m$. Then $f(X)$ is a compact subset of $\mathbb{R}^m$.*

COROLLARY 6. *Suppose that $f : X \to Y$ is a continuous function from a compact subset X of $\mathbb{R}^n$ to a subset Y of $\mathbb{R}^m$. If f is both* one-to-one *and* onto, *then the inverse function $f^{-1} : Y \to X$ defined by*

$$f^{-1}(y) = x \text{ where } x \text{ is the unique point in } X \text{ satisfying } f(x) = y,$$

is a continuous map.

Proof (of Corollary 5): If $\{G_\alpha\}_{\alpha \in A}$ is an open cover of $f(X)$, then $\left\{f^{-1}(G_\alpha)\right\}_{\alpha \in A}$ is an open cover of X, hence has a finite subcover $\left\{f^{-1}(G_{\alpha_k})\right\}_{k=1}^N$. But then $\{G_{\alpha_k}\}_{k=1}^N$ is a finite subcover of $f(X)$ since

$$f(X) \subset f\left(\bigcup_{k=1}^N f^{-1}(G_{\alpha_k})\right) \subset \bigcup_{k=1}^N f\left(f^{-1}(G_{\alpha_k})\right) \subset \bigcup_{k=1}^N G_{\alpha_k}.$$

Note that it is *not* in general true that $f^{-1}(f(G)) \subset G$.

Proof (of Corollary 6): Let G be an open subset of X. We must show that $\left(f^{-1}\right)^{-1}(G)$ is open in Y. Note that since f is one-to-one and onto, we have $\left(f^{-1}\right)^{-1}(G) = f(G)$. Now $G^c = X \setminus G$ is closed in X, hence compact, and so Corollary 5 shows that $f(G^c)$ is compact, hence closed in Y, so $f(G^c)^c$ is open in Y. But again using that f is one-to-one and onto shows that $f(G) = f(G^c)^c$, and so we are done.

REMARK 4. *Compactness is essential in this corollary since the map*

$$f : [0, 2\pi) \to \mathbb{T} \equiv \{z \in \mathbb{C} : |z| = 1\} \text{ defined by } f(\theta) = e^{i\theta} = (\cos\theta, \sin\theta),$$

and takes $[0, 2\pi)$ one-to-one and continuously onto $\mathbb{T}$, yet the inverse map fails to be continuous at $z = 1$. Indeed, for points z on the circle just below 1, *$f^{-1}(z)$ is close to 2π, while $f^{-1}(1) = 0$.*

Proof (of Theorem 9): Suppose first that f is continuous on X. We must show that (2.4) holds. So let G be an open subset of Y. We must now show that for every $p \in f^{-1}(G)$ there is $r > 0$ (depending on p) such that $B(p, r) \subset f^{-1}(G)$. Fix $p \in f^{-1}(G)$. Since G is open and $f(p) \in G$ we can pick $\varepsilon > 0$ such that $B(f(p), \varepsilon) \subset G$. But then by the continuity of f there is $\delta > 0$ such that (2.2) holds, i.e. $f(B(p, \delta)) \subset B(f(p), \varepsilon) \subset G$. It follows that

$$B(p, \delta) \subset f^{-1}(f(B(p, \delta))) \subset f^{-1}(G).$$

Conversely suppose that (2.4) holds. We must show that f is continuous at every $p \in X$. So fix $p \in X$. We must now show that for every $\varepsilon > 0$ there is $\delta > 0$ such that (2.2) holds, i.e. $f(B(p, \delta)) \subset B(f(p), \varepsilon)$. Fix $\varepsilon > 0$. Since $B(f(p), \varepsilon)$ is open, we have that $f^{-1}(B(f(p), \varepsilon))$ is open by (2.4). Since $p \in f^{-1}(B(f(p), \varepsilon))$ there is thus $\delta > 0$ such that $B(p, \delta) \subset f^{-1}(B(f(p), \varepsilon))$. It follows that

$$f(B(p, \delta)) \subset f\left(f^{-1}(B(f(p), \varepsilon))\right) \subset B(f(p), \varepsilon).$$

We now show that continuity is stable under composition of maps.

THEOREM 10. *Suppose that X, Y, Z are subsets of $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^p$ respectively. If $f : X \to Y$ and $g : Y \to Z$ are both continuous maps, then so is the composition $h = g \circ f : X \to Z$ defined by*

$$h(x) = g(f(x)), \qquad x \in X.$$

Proof: If G is open in Z, then

$$h^{-1}(G) = f^{-1}\left(g^{-1}(G)\right)$$

is open since g continuous implies $g^{-1}(G)$ is open by Theorem 9, and then f continuous implies $f^{-1}\left(g^{-1}(G)\right)$ is open by Theorem 9. Thus h is continuous by Theorem 9.

Continuity at a point is also easily handled using Definition 4. We leave the proof of the following theorem to the reader.

THEOREM 11. *Suppose that* X, Y, Z *are subsets of* $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^p$ *respectively. If* $p \in X$ *and* $f : X \to Y$ *is continuous at* p *and* $g : f(X) \to Z$ *is continuous at* $f(p)$, *then the composition* $h = g \circ f : X \to Z$ *is continuous at* p.

2.1. Real and complex-valued continuous functions. Here is an elementary consequence of the familiar limit theorems for sums, differences, products and quotients of complex-valued functions.

PROPOSITION 1. *If* f *and* g *are continuous complex-valued functions on a subset* X *of* $\mathbb{R}^n$, *then so are the functions* $f + g$ *and* fg. *If in addition* g *never vanishes, then* $\frac{f}{g}$ *is also continuous on* X.

Here is an extremely useful consequence of Corollary 5 when the target space Y is the real numbers.

THEOREM 12. *Suppose that* X *is a compact subset of* $\mathbb{R}^n$ *and* $f : X \to \mathbb{R}$ *is continuous. Then there exist points* $p, q \in X$ *satisfying*

$$f(p) = \sup f(X) \text{ and } f(q) = \inf f(X).$$

REMARK 5. *Compactness of* X *is essential here as evidenced by the following example. If* X *is the open interval* $(0,1)$ *and* $f : (0,1) \to (0,1)$ *is the identity map defined by* $f(x) = x$, *then* f *is continuous and*

$$\begin{aligned} \sup f((0,1)) &= \sup(0,1) = 1, \\ \inf f((0,1)) &= \inf(0,1) = 0. \end{aligned}$$

However, there are no points $p, q \in (0,1)$ *satisfying either* $f(p) = 1$ *or* $f(q) = 0$.

Proof (of Theorem 12): Corollary 5 shows that $f(X)$ is compact. Lemmas 1 and 2 now show that $f(X)$ is a closed and bounded subset of $\mathbb{R}$. Thus $\sup f(X)$ exists and $\sup f(X) \in f(X)$, i.e. there is $p \in X$ such that $\sup f(X) = f(p)$. Similarly there is $q \in X$ satisfying $\inf f(X) = f(q)$.

Now consider a complex-valued function $f : X \to \mathbb{C}$ on a subset X of $\mathbb{R}^n$, and let $u : X \to \mathbb{R}$ and $v : X \to \mathbb{R}$ be the real and imaginary parts of f defined by

$$\begin{aligned} u(x) &= \operatorname{Re} f(x) \equiv \frac{f(x) + \overline{f(x)}}{2}, \\ v(x) &= \operatorname{Im} f(x) \equiv \frac{f(x) - \overline{f(x)}}{2i}, \end{aligned}$$

for $x \in X$. It is easy to see that f is continuous at a point $p \in X$ if and only if each of u and v is continuous at p. Indeed, the inequalities

$$\max\{|a|, |b|\} \leq \sqrt{|a|^2 + |b|^2} \leq |a| + |b|$$

show that if (2.1) holds for f (with $E = X$), i.e.

$$d_{\mathbb{C}}(f(x), f(p)) < \varepsilon \text{ whenever } d_X(x, p) < \delta,$$

then it also holds with f replaced by u or by v:

$$\begin{aligned} d_{\mathbb{R}}(u(x), u(p)) &= |u(x) - u(p)| \\ &\leq \sqrt{|u(x) - u(p)|^2 + |v(x) - v(p)|^2} \\ &= d_{\mathbb{C}}(f(x), f(p)) < \varepsilon \\ &\quad \text{whenever } d_X(x, p) < \delta. \end{aligned}$$

Similarly, if (2.1) holds for both u and v then it holds for f but with ε replaced by 2ε:

$$\begin{aligned} d_{\mathbb{C}}(f(x), f(p)) &= \sqrt{|u(x) - u(p)|^2 + |v(x) - v(p)|^2} \\ &\leq |u(x) - u(p)| + |v(x) - v(p)| \\ &= d_{\mathbb{R}}(u(x), u(p)) + d_{\mathbb{R}}(v(x), v(p)) < 2\varepsilon \\ &\quad \text{whenever } d_X(x, p) < \delta. \end{aligned}$$

The same considerations apply equally well to Euclidean space $\mathbb{R}^m$ (recall that $\mathbb{C} = \mathbb{R}^2$ as metric spaces) and we have the following theorem. Recall that the dot product of two vectors $\mathbf{z} = (z_1, ..., z_m)$ and $\mathbf{w} = (w_1, ..., w_m)$ in $\mathbb{R}^m$ is given by $\mathbf{z} \cdot \mathbf{w} = \sum_{k=1}^m z_k w_k$.

Theorem 13. *Let X be a subset of $\mathbb{R}^n$ and suppose $\mathbf{f} : X \to \mathbb{R}^m$. Let $f_k(x)$ be the component functions defined by $\mathbf{f}(x) = (f_1(x), ..., f_m(x))$ for $1 \leq k \leq m$.*

(1) *The vector-valued function $\mathbf{f} : X \to \mathbb{R}^m$ is continuous at a point $p \in X$* if and only if *each component function $f_k : X \to \mathbb{R}$ is continuous at p.*
(2) *If both $\mathbf{f} : X \to \mathbb{R}^m$ and $\mathbf{g} : X \to \mathbb{R}^m$ are continuous at p then so are $\mathbf{f} + \mathbf{g} : X \to \mathbb{R}^m$ and $\mathbf{f} \cdot \mathbf{g} : X \to \mathbb{R}$.*

Here are some simple facts associated with the component functions on Euclidean space.

- For each $1 \leq j \leq n$, the component function $\mathbf{w} = (w_1, ..., w_n) \to w_j$ is continuous from $\mathbb{R}^n$ to $\mathbb{R}$.
- The length function $\mathbf{w} = (w_1, ..., w_n) \to |\mathbf{w}|$ is continuous from $\mathbb{R}^n$ to $[0, \infty)$; in fact we have the so-called reverse triangle inequality:
$$||\mathbf{z}| - |\mathbf{w}|| \leq |\mathbf{z} - \mathbf{w}|, \qquad \mathbf{z}, \mathbf{w} \in \mathbb{R}^n.$$
- Every monomial function $\mathbf{w} = (w_1, ..., w_n) \to w_1^{k_1} w_2^{k_2} ... w_n^{k_n}$ is continuous from $\mathbb{R}^n$ to $\mathbb{R}$.
- Every polynomial $P(\mathbf{w}) = \sum_{k_1 + ... k_n \leq N} a_{k_1, ... k_n} w_1^{k_1} w_2^{k_2} ... w_n^{k_n}$ is continuous from $\mathbb{R}^n$ to $\mathbb{R}$.

3. Uniform continuity

A function $f : X \to Y$ that is continuous from a subset X of $\mathbb{R}^n$ to another subset Y of $\mathbb{R}^m$ satisfies Definition 4 at each point p in X, namely for every $p \in X$ and $\varepsilon > 0$ there is $\delta_p > 0$ (note the dependence on p) such that (2.1) holds with $E = X$:

$$d_Y(f(x), f(p)) < \varepsilon \text{ whenever } d_X(x, p) < \delta_p. \tag{3.1}$$

In general we cannot choose $\delta > 0$ to be independent of p. For example, the function $f(x) = \frac{1}{x}$ is continuous on the open interval $(0,1)$, but if we want

$$\varepsilon > d_Y(f(x), f(p)) = \left|\frac{1}{x} - \frac{1}{p}\right| \text{ whenever } |p - x| < \delta,$$

we cannot take $p = \delta$ since then x could be arbitrarily close to 0, and so $\frac{1}{x}$ could be arbitrarily large. In this example, $X = (0,1)$ is not compact and this turns out to be the reason we cannot choose $\delta > 0$ to be independent of p. The surprising property that continuous functions f on a *compact* metric space X have is that we *can* indeed choose $\delta > 0$ to be independent of p in (3.1). We first give a name to this surprising property; we call it *uniform continuity* on X.

Definition 5. *Suppose that $f : X \to Y$ maps a subset X of $\mathbb{R}^n$ into a subset Y of $\mathbb{R}^m$. We say that f is* uniformly continuous *on X if for every $\varepsilon > 0$ there is $\delta > 0$ such that*

$$d_Y(f(x), f(p)) < \varepsilon \textit{ whenever } d_X(x,p) < \delta.$$

The next theorem plays a crucial role in the theory of integration.

Theorem 14. *Suppose that $f : X \to Y$ is a continuous map from a* compact *subset X of $\mathbb{R}^n$ into a subset Y of $\mathbb{R}^m$. Then f is uniformly continuous on X.*

Proof: Suppose $\varepsilon > 0$. Since f is continous on X, (2.2) shows that for each point $p \in X$, there is $\delta_p > 0$ such that

$$f(B(p, \delta_p)) \subset B\left(f(p), \frac{\varepsilon}{2}\right). \tag{3.2}$$

Since X is compact, the open cover $\left\{B\left(p, \frac{\delta_p}{2}\right)\right\}_{p \in X}$ has a finite subcover $\left\{B\left(p_k, \frac{\delta_{p_k}}{2}\right)\right\}_{k=1}^N$. Now define

$$\delta = \min\left\{\frac{\delta_{p_k}}{2}\right\}_{k=1}^N.$$

Since the minimum is taken over *finitely* many positive numbers (thanks to the *finite* subcover, which in turn owes its existence to the *compactness* of X), we have $\delta > 0$.

Now suppose that $x, p \in X$ satisfy $d_X(x,p) < \delta$. We will show that

$$d_Y(f(x), f(p)) < \varepsilon.$$

Choose k so that $p \in B\left(p_k, \frac{\delta_{p_k}}{2}\right)$. Then we have using the triangle inequality in X that

$$d_X(x, p_k) \le d_X(x,p) + d_X(p, p_k) < \delta + \frac{\delta_{p_k}}{2} \le \frac{\delta_{p_k}}{2} + \frac{\delta_{p_k}}{2} = \delta_{p_k},$$

so that both p and x lie in the ball $B(p_k, \delta_{p_k})$. It follows from (3.2) that both $f(p)$ and $f(x)$ lie in

$$f(B(p_k, \delta_{p_k})) \subset B\left(f(p_k), \frac{\varepsilon}{2}\right).$$

Finally an application of the triangle inequality in Y shows that

$$d_Y(f(x), f(p)) \le d_Y(f(x), f(p_k)) + d_Y(f(p_k), f(p)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

4. Connectedness

Definition 6. *A subset X of $\mathbb{R}^n$ is said to be* connected *if it is* not *possible to write $X = E \dot{\cup} F$ where E and F are disjoint nonempty relatively open subsets of X. A set that is not connected is said to be* disconnected.

Equivalently, X is disconnected if it has a nonempty proper relatively *clopen* subset (a relatively clopen subset of X is one that is simultaneously relatively open and relatively closed in X).

Lemma 4. *A subset X of $\mathbb{R}^n$ is disconnected if and only if there are nonempty subsets E and F of $\mathbb{R}^n$ with $X = E \dot{\cup} F$ and*

$$\overline{E} \cap F = \emptyset \text{ and } E \cap \overline{F} = \emptyset, \tag{4.1}$$

where the closures refer to the Euclidean space $\mathbb{R}^n$.

Proof: It is not hard to see that E is a relatively open subset of X if and only if $E \cap \overline{F} = \emptyset$ where $F = X \setminus E$. Similarly, F is relatively open in X if and only if $\overline{E} \cap F = \emptyset$. Finally, E is relatively clopen in X if and only if both E and $F = X \setminus E$ are relatively open in X.

The connected subsets of the real line are especially simple - they are precisely the intervals

$$[a, b], (a, b), [a, b), (a, b]$$

lying in $\mathbb{R}$ with $-\infty \leq a \leq b \leq \infty$ (we do not consider any case where a or b is $\pm\infty$ and lies next to either [or]).

Theorem 15. *The connected subsets of the real numbers $\mathbb{R}$ are precisely the intervals.*

Proof: Consider first a nonempty connected subset Y of $\mathbb{R}$. If $a, b \in Y$, and $a < c < b$, then we must also have $c \in Y$ since otherwise $Y \cap (-\infty, c)$ is clopen in Y. Thus the set Y has the *intermediate value property* ($a, b \in Y$ and $a < c < b$ implies $c \in Y$), and it is now easy to see using the Least Upper Bound Property of $\mathbb{R}$, that Y is an interval. Conversely, if Y is a disconnected subset of $\mathbb{R}$, then Y has a nonempty proper clopen subset E. We can then find two points $a, b \in Y$ with $a \in E$ and $b \in F \equiv Y \setminus E$ and (without loss of generality) $a < b$. Set

$$c \equiv \sup (E \cap [a, b]) .$$

Then we have $c \in \overline{E}$, and so $c \notin F$ by (4.1). If also $c \notin E$, then Y fails the intermediate value property and so cannot be an interval. On the other hand, if $c \in E$ then $c \notin \overline{F}$ (the *closure* of F), and so there is $d \in (c, b) \setminus F$. But then $d \notin E$ since $d > c$ and so lies in $(a, b) \setminus Y$, which again shows that Y fails the intermediate value property and so cannot be an interval.

Connected sets behave the same way as compact sets under pushforward by a continuous map.

Theorem 16. *Suppose $f : X \to Y$ is a continuous map from a subset X of $\mathbb{R}^n$ to a subset Y of $\mathbb{R}^m$. If X is connected, then $f(X)$ is connected.*

Proof: We may assume that $Y = f(X)$. If Y is disconnected, there are disjoint nonempty open subsets E and F with $Y = E \dot{\cup} F$. But then $X = f^{-1}(E) \dot{\cup} f^{-1}(F)$ where both $f^{-1}(E)$ and $f^{-1}(F)$ are open in X by Theorem 9. This shows that X is disconnected as well, and completes the proof of the (contrapositive of the) theorem.

COROLLARY 7. *If $f : \mathbb{R} \to \mathbb{R}$ is continuous, then f takes intervals to intervals, and in particular, f takes closed bounded intervals to closed bounded intervals.*

Note that this corollary yields two familiar theorems from first year calculus, the Intermediate Value Theorem (real continuous functions on an interval attain their intermediate values) and the Extreme Value Theorem (real continuous functions on a closed bounded interval attain their extreme values).

Proof: Apply Theorems 16, 5 and 5.

Finally we have the following simple description of open subsets of the real numbers.

PROPOSITION 2. *Every open subset G of the real numbers $\mathbb{R}$ can be uniquely written as an at most countable pairwise disjoint union of open intervals $\{I_n\}_{n\geq 1}$:*

$$G = \dot{\bigcup_{n\geq 1}} I_n.$$

Proof: For $x \in G$ let

$$I_x = \bigcup \{\text{all open intervals containing } x \text{ that are contained in } G\}.$$

It is easy to see that I_x is an open interval and that if $x, y \in G$ then

$$\text{either } I_x = I_y \text{ or } I_x \cap I_y = \emptyset.$$

This shows that G is a union $\bigcup_{\alpha\in A} I_\alpha$ of pairwise disjoint open intervals. To see that this union is at most countable, simply pick a rational number r_α in each I_α. The uniqueness is left as an exercise for the reader.

Part 2

Integration and differentiation

In the second part of these notes we begin with the problem of describing the inverse operation to that of differentiation, commonly called *integration*. There are four widely recognized theories of integration:

- Riemann integration - the workhorse of integration theory that provides us with the most basic form of the fundamental theorem of calculus;
- Riemann-Stieltjes integration - that extends the idea of integrating the infinitesmal dx to that of the more general infinitesmal $d\alpha(x)$ for an increasing function α.
- Lebesgue integration - that overcomes a shortcoming of the Riemann theory by permitting a robust theory of limits of functions, all at the expense of a complicated theory of 'measure' of a set.
- Henstock-Kurtzweil integration - that includes the Riemann and Lebesgue theories and has the advantages that it is quite similar in spirit to the intuitive Riemann theory, and avoids much of the complication of measurability of sets in the Lebesgue theory. However, it has the drawback of limited scope for generalization.

In Chapter 3 we follow Rudin [**3**] and use uniform continuity to develop the standard theory of the Riemann and Riemann-Stieltjes integrals. A short detour is taken to introduce the more powerful Henstock-Kurtzweil integral, and we use compactness to prove its uniqueness and extension properties.

Chapter 4 draws on Stein and Shakarchi [**6**] to provide a rapid and transparent introduction to the theory of the Lebesgue integral on the real line.

Chapter 5 proves the Banach-Tarski paradox by exploiting the existence of a free nonabelian group of rank 2 in the rotation group SO_3 in three dimensions. There is no better advertisement for resticting matters to measurable sets.

Chapter 6 uses Urysohn's Lemma to establish the Riesz representation theorem on locally compact Hausdorff spaces, and constructs Lebesgue measure on Euclidean spaces. Regularity of measures is treated in some detail, and the Tietze extension theorem is used to prove Lusin's theorem.

Chapter 7 introduces the Lebesgue spaces $L^p(\mu)$ and develops their elementary theory including duality theory. The Baire category theorem is used to prove the classical consequences in the more general setting of Banach spaces, namely the uniform boundedness principle, the open mapping theorem and the closed graph theorem, together with some applications. The special case $p = 2$ is further developed in the context of Hilbert spaces.

Chapter 8 introduces complex measures and proves the Radon-Nikodym theorem using Hilbert space theory.

Chapter 9 discusses differentiation of integrals using shifted dyadic grids.

Chapter 10 introduces integration on product spaces and proves Fubini's theorem.

CHAPTER 3

Riemann and Riemann-Stieltjes integration

Let $f : [0,1] \to \mathbb{R}$ be a bounded function on the closed unit interval $[0,1]$. In *Riemann's* theory of integration, we partition the *domain* $[0,1]$ of the function into finitely many disjoint subintervals

$$[0,1] = \bigcup_{n=1}^{N} [x_{n-1}, x_n],$$

and denote the partition by $\mathcal{P} = \{0 = x_0 < x_1 < ... < x_N = 1\}$ and the length of the subinterval $[x_{n-1}, x_n]$ by $\triangle x_n = x_n - x_{n-1} > 0$. Then we define *upper and lower Riemann sums* associated with the partition $\mathcal{P}$ by

$$\begin{aligned} U(f;\mathcal{P}) &= \sum_{n=1}^{N} \left(\sup_{[x_{n-1},x_n]} f \right) \triangle x_n, \\ L(f;\mathcal{P}) &= \sum_{n=1}^{N} \left(\inf_{[x_{n-1},x_n]} f \right) \triangle x_n. \end{aligned}$$

Note that the suprema and infima are finite since f is bounded by assumption. Next we define the *upper and lower Riemann integrals* of f on $[0,1]$ by

$$\mathcal{U}(f) = \inf_{\mathcal{P}} U(f;\mathcal{P}), \qquad \mathcal{L}(f) = \sup_{\mathcal{P}} L(f;\mathcal{P}).$$

Thus the upper Riemann integral $\mathcal{U}(f)$ is the "smallest" of all the upper sums, and the lower Riemann integral is the "largest" of all the lower sums.

We can show that any upper sum is always larger than any lower sum by considering the *refinement* of two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$: $\mathcal{P}_1 \cup \mathcal{P}_2$ denotes the paritition whose points consist of the union of the points in $\mathcal{P}_1$ and $\mathcal{P}_2$ and ordered to be strictly increasing.

LEMMA 5. *Suppose $f : [0,1] \to \mathbb{R}$ is bounded. If $\mathcal{P}_1$ and $\mathcal{P}_2$ are any two partitions of $[0,1]$, then*

$$U(f;\mathcal{P}_1) \geq U(f;\mathcal{P}_1 \cup \mathcal{P}_2) \geq L(f;\mathcal{P}_1 \cup \mathcal{P}_2) \geq L(f;\mathcal{P}_2). \tag{0.2}$$

Proof: Let

$$\begin{aligned} \mathcal{P}_1 &= \{0 = x_0 < x_1 < ... < x_M = 1\}, \\ \mathcal{P}_2 &= \{0 = y_0 < y_1 < ... < y_N = 1\}, \\ \mathcal{P}_1 \cup \mathcal{P}_2 &= \{0 = z_0 < z_1 < ... < z_P = 1\}. \end{aligned}$$

Fix a subinterval $[x_{n-1}, x_n]$ of the partition $\mathcal{P}_1$. Suppose that $[x_{n-1}, x_n]$ contains exactly the following increasing sequence of points in the partition $\mathcal{P}_1 \cup \mathcal{P}_2$:

$$z_{\ell_n} < z_{\ell_n+1} < ... < z_{\ell_n+m_n},$$

i.e. $z_{\ell_n} = x_{n-1}$ and $z_{\ell_n+m_n} = x_n$. Then we have

$$\begin{aligned}\left(\sup_{[x_{n-1},x_n]} f\right) \triangle x_n &= \left(\sup_{[x_{n-1},x_n]} f\right)\left(\sum_{j=1}^{m_n} \triangle z_{\ell_n+j}\right) \\ &\geq \sum_{j=1}^{m_n}\left(\sup_{[z_{\ell_n+j-1},z_{\ell_n+j}]} f\right) \triangle z_{\ell_n+j},\end{aligned}$$

since $\sup_{[z_{\ell_n+j-1},z_{\ell_n+j}]} f \leq \sup_{[x_{n-1},x_n]} f$ when $[z_{\ell_n+j-1}, z_{\ell_n+j}] \subset [x_{n-1}, x_n]$. If we now sum over $1 \leq n \leq M$ we get

$$\begin{aligned}U(f;\mathcal{P}_1) &= \sum_{n=1}^{M}\left(\sup_{[x_{n-1},x_n]} f\right) \triangle x_n \\ &\geq \sum_{n=1}^{M}\sum_{j=1}^{m_n}\left(\sup_{[z_{\ell_n+j-1},z_{\ell_n+j}]} f\right) \triangle z_{\ell_n+j} \\ &= \sum_{p=1}^{P}\left(\sup_{[z_{p-1},z_p]} f\right) \triangle z_p = U(f;\mathcal{P}_1 \cup \mathcal{P}_2).\end{aligned}$$

Similarly we can prove that

$$L(f;\mathcal{P}_2) \leq L(f;\mathcal{P}_1 \cup \mathcal{P}_2).$$

Since we trivially have $L(f;\mathcal{P}_1 \cup \mathcal{P}_2) \leq U(f;\mathcal{P}_1 \cup \mathcal{P}_2)$, the proof of the lemma is complete.

Now in (0.2) take the infimum over $\mathcal{P}_1$ and the supremum over $\mathcal{P}_2$ to obtain that

$$\mathcal{U}(f) \geq \mathcal{L}(f),$$

which says that the *upper* Riemann integral of f is always equal to or greater than the *lower* Riemann integral of f. Finally we say that f is Riemann integrable on $[0,1]$, written $f \in \mathcal{R}[0,1]$, if $\mathcal{U}(f) = \mathcal{L}(f)$, and we denote the common value by $\int_0^1 f$ or $\int_0^1 f(x)\,dx$.

We can of course repeat this line of definition and reasoning for any bounded closed interval $[a,b]$ in place of the closed unit interval $[0,1]$. We summarize matters in the following definition.

Definition 7. *Let* $f : [a,b] \to \mathbb{R}$ *be a bounded function. For any partition* $\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$ *of* $[a,b]$ *we define* upper and lower Riemann sums *by*

$$\begin{aligned}U(f;\mathcal{P}) &= \sum_{n=1}^{N}\left(\sup_{[x_{n-1},x_n]} f\right) \triangle x_n, \\ L(f;\mathcal{P}) &= \sum_{n=1}^{N}\left(\inf_{[x_{n-1},x_n]} f\right) \triangle x_n.\end{aligned}$$

Set

$$\mathcal{U}(f) = \inf_{\mathcal{P}} U(f;\mathcal{P}), \qquad \mathcal{L}(f) = \sup_{\mathcal{P}} L(f;\mathcal{P}),$$

where the infimum and supremum are taken over all partitions $\mathcal{P}$ of $[a,b]$. We say that f is Riemann integrable *on $[a,b]$, written $f \in \mathcal{R}[a,b]$, if $\mathcal{U}(f) = \mathcal{L}(f)$, and we denote the common value by*

$$\int_a^b f \quad \text{or} \quad \int_a^b f(x)\,dx.$$

A more substantial generalization of the line of definition and reasoning above can be obtained on a closed interval $[a,b]$ by considering in place of the positive quantities $\triangle x_n = x_n - x_{n-1}$ associated with a partition

$$\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$$

of $[a,b]$, the more general nonnegative quantities

$$\triangle\alpha_n = \alpha(x_n) - \alpha(x_{n-1}), \qquad 1 \leq n \leq N,$$

where $\alpha : [a,b] \to \mathbb{R}$ is *nondecreasing*. This leads to the notion of the Riemann-Stieltjes integral associated with a nondecreasing function $\alpha : [a,b] \to \mathbb{R}$.

Definition 8. *Let $f : [a,b] \to \mathbb{R}$ be a bounded function and suppose $\alpha : [a,b] \to \mathbb{R}$ is nondecreasing. For any partition $\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$ of $[a,b]$ we define* upper and lower Riemann sums *by*

$$\begin{aligned} U(f;\mathcal{P},\alpha) &= \sum_{n=1}^{N} \left(\sup_{[x_{n-1},x_n]} f \right) \triangle\,\alpha_n, \\ L(f;\mathcal{P},\alpha) &= \sum_{n=1}^{N} \left(\inf_{[x_{n-1},x_n]} f \right) \triangle\,\alpha_n. \end{aligned}$$

Set

$$\mathcal{U}(f,\alpha) = \inf_{\mathcal{P}} U(f;\mathcal{P},\alpha), \qquad \mathcal{L}(f,\alpha) = \sup_{\mathcal{P}} L(f;\mathcal{P},\alpha),$$

where the infimum and supremum are taken over all partitions $\mathcal{P}$ of $[a,b]$. We say that f is Riemann-Stieltjes integrable *on $[a,b]$, written $f \in \mathcal{R}_\alpha[a,b]$, if $\mathcal{U}(f,\alpha) = \mathcal{L}(f,\alpha)$, and we denote the common value by*

$$\int_a^b f d\alpha \quad \text{or} \quad \int_a^b f(x)\,d\alpha(x).$$

The lemma on partitions above generalizes immediately to the setting of the Riemann-Stieltjes integral.

Lemma 6. *Suppose $f : [a,b] \to \mathbb{R}$ is bounded and $\alpha : [a,b] \to \mathbb{R}$ is nondecreasing. If $\mathcal{P}_1$ and $\mathcal{P}_2$ are any two partitions of $[a,b]$, then*

$$U(f;\mathcal{P}_1,\alpha) \geq U(f;\mathcal{P}_1 \cup \mathcal{P}_2,\alpha) \geq L(f;\mathcal{P}_1 \cup \mathcal{P}_2,\alpha) \geq L(f;\mathcal{P}_2,\alpha). \tag{0.3}$$

0.1. Existence of the Riemann-Stieltjes integral. The difficult question now arises as to exactly which bounded functions f are Riemann-Stieltjes integrable with respect to a given nondecreasing α on $[a,b]$. We will content ourselves with showing two results. Suppose f is bounded on $[a,b]$ and α is nondecreasing on $[a,b]$. Then

- $f \in \mathcal{R}_\alpha[a,b]$ if in addition f is *continuous* on $[a,b]$;
- $f \in \mathcal{R}_\alpha[a,b]$ if in addition f is *monotonic* on $[a,b]$ and α is *continuous* on $[a,b]$.

Both proofs will use the Cauchy criterion for existence of the integral $\int_a^b f d\alpha$ when $f : [a, b] \to \mathbb{R}$ is bounded and $\alpha : [a, b] \to \mathbb{R}$ is nondecreasing:

$$\text{For every } \varepsilon > 0 \text{ there is a partition } \mathcal{P} \text{ of } [a,b] \text{ such that } U(f;\mathcal{P},\alpha) - L(f;\mathcal{P},\alpha) < \varepsilon. \tag{0.4}$$

Clearly, if (0.4) holds, then from (0.3) we obtain that for each $\varepsilon > 0$ that there is a partition $\mathcal{P}_\varepsilon$ satisfying

$$\begin{aligned} \mathcal{U}(f,\alpha) - \mathcal{L}(f,\alpha) &= \inf_{\mathcal{P}} U(f;\mathcal{P},\alpha) - \sup_{\mathcal{P}} L(f;\mathcal{P},\alpha) \\ &\leq U(f;\mathcal{P}_\varepsilon,\alpha) - L(f;\mathcal{P}_\varepsilon,\alpha) < \varepsilon. \end{aligned}$$

It follows that $\mathcal{U}(f,\alpha) = \mathcal{L}(f,\alpha)$ and so $\int_a^b f d\alpha$ exists. Conversely, given $\varepsilon > 0$ there are partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ satisfying

$$\begin{aligned} \mathcal{U}(f,\alpha) &= \inf_{\mathcal{P}} U(f;\mathcal{P},\alpha) > U(f;\mathcal{P}_1,\alpha) - \frac{\varepsilon}{2}, \\ \mathcal{L}(f,\alpha) &= \sup_{\mathcal{P}} L(f;\mathcal{P},\alpha) < L(f;\mathcal{P}_2,\alpha) + \frac{\varepsilon}{2}. \end{aligned}$$

Inequality (0.3) now shows that

$$\begin{aligned} U(f;\mathcal{P}_1\cup\mathcal{P}_2,\alpha) - L(f;\mathcal{P}_1\cup\mathcal{P}_2,\alpha) &\leq U(f;\mathcal{P}_1,\alpha) - L(f;\mathcal{P}_2,\alpha) \\ &< \left(\mathcal{U}(f,\alpha) + \frac{\varepsilon}{2}\right) - \left(\mathcal{L}(f,\alpha) - \frac{\varepsilon}{2}\right) = \varepsilon \end{aligned}$$

since $\mathcal{U}(f,\alpha) = \mathcal{L}(f,\alpha)$ if $\int_a^b f d\alpha$ exists. Thus we can take $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2$ in (0.4).

The existence of $\int_a^b f d\alpha$ when f is continuous will use Theorem 14 on uniform continuity in a crucial way.

THEOREM 17. *Suppose that $f : [a,b] \to \mathbb{R}$ is continuous and $\alpha : [a,b] \to \mathbb{R}$ is nondecreasing. Then $f \in \mathcal{R}_\alpha[a,b]$.*

Proof: We will show that the Cauchy criterion (0.4) holds. Fix $\varepsilon > 0$. By Theorem 14 f is uniformly continuous on the compact set $[a,b]$, so there is $\delta > 0$ such that

$$|f(x) - f(x')| \leq \frac{\varepsilon}{\alpha(b) - \alpha(a)} \text{ whenever } |x - x'| \leq \delta.$$

Let $\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$ be any partition of $[a,b]$ for which

$$\max_{1\leq n\leq N} \triangle x_n < \delta.$$

Then we have

$$\sup_{[x_{n-1},x_n]} f - \inf_{[x_{n-1},x_n]} f \leq \sup_{x,x'\in[x_{n-1},x_n]} |f(x) - f(x')| \leq \varepsilon,$$

since $|x - x'| \leq \triangle x_n < \delta$ when $x, x' \in [x_{n-1}, x_n]$ by our choice of $\mathcal{P}$. Now we compute that

$$\begin{aligned} U(f;\mathcal{P},\alpha) - L(f;\mathcal{P},\alpha) &= \sum_{n=1}^N \left(\sup_{[x_{n-1},x_n]} f - \inf_{[x_{n-1},x_n]} f\right) \triangle\alpha_n \\ &\leq \sum_{n=1}^N \left(\frac{\varepsilon}{\alpha(b) - \alpha(a)}\right) \triangle\alpha_n = \varepsilon, \end{aligned}$$

which is (0.4) as required.

REMARK 6. *Observe that it makes no logical difference if we replace strict inequality* $<$ *with* $\leq$ *in '*$\varepsilon-\delta$ *type' definitions. We have used this observation twice in the above proof, and will continue to use it without further comment in the sequel.*

The proof of the next existence result uses the intermediate value theorem for continuous functions.

THEOREM 18. *Suppose that* $f:[a,b]\rightarrow\mathbb{R}$ *is monotone and* $\alpha:[a,b]\rightarrow\mathbb{R}$ *is nondecreasing and continuous. Then* $f\in\mathcal{R}_{\alpha}[a,b]$.

Proof: We will show that the Cauchy criterion (0.4) holds. Fix $\varepsilon>0$ and suppose without loss of generality that f is nondecreasing on $[a,b]$. Let $N\geq 2$ be a positive integer. Since α is continuous we can use the intermediate value theorem to find points $x_n\in(a,b)$ such that $x_0=a$, $x_N=b$ and

$$\alpha(x_n)=\alpha(a)+\frac{n}{N}(\alpha(b)-\alpha(a)),\qquad 1\leq n\leq N-1.$$

Since α is nondecreasing we have $x_{n-1}<x_n$ for all $1\leq n\leq N$, and it follows that

$$\mathcal{P}=\{a=x_0<x_1<...<x_N=b\}$$

is a partition of $[a,b]$ satisfying

$$\triangle\alpha_n=\alpha(x_n)-\alpha(x_{n-1})=\frac{\alpha(b)-\alpha(a)}{N}<\frac{\varepsilon}{f(b)-f(a)},$$

provided we take N large enough. With such a partition $\mathcal{P}$ we compute

$$\begin{aligned}U(f;\mathcal{P},\alpha)-L(f;\mathcal{P},\alpha) &= \sum_{n=1}^{N}\left(\sup_{[x_{n-1},x_n]}f-\inf_{[x_{n-1},x_n]}f\right)\triangle\alpha_n\\ &\leq \frac{\varepsilon}{f(b)-f(a)}\sum_{n=1}^{N}\left(\sup_{[x_{n-1},x_n]}f-\inf_{[x_{n-1},x_n]}f\right)\\ &= \frac{\varepsilon}{f(b)-f(a)}\sum_{n=1}^{N}(f(x_n)-f(x_{n-1}))=\varepsilon,\end{aligned}$$

This proves (0.4) as required.

0.2. A stronger form of the definition of the Riemann integral. For the Riemann integral there is another formulation of the definition of $\int_a^b f$ that appears at first sight to be much stronger (and which doesn't work for general nondecreasing α in the Riemann-Stieltjes integral). For any partition $\mathcal{P}=\{a=x_0<x_1<...<x_N=b\}$, set $\|\mathcal{P}\|=\max_{1\leq n\leq N}\triangle x_n$, called the *norm* of $\mathcal{P}$. Now if $\int_a^b f$ exists, then for every $\varepsilon>0$ there is by the Cauchy criterion (0.4) a partition $\mathcal{P}=\{a=x_0<x_1<...<x_N=b\}$ such that

$$U(f;\mathcal{P})-L(f;\mathcal{P})<\frac{\varepsilon}{2}.$$

Now define δ to be the smaller of the two positive numbers

$$\min_{1\leq n\leq N}\triangle x_n \text{ and } \frac{\varepsilon}{2N\ diam\ f([a,b])}.$$

CLAIM 1. *If* $\mathcal{Q}=\{a=y_0<y_1<...<y_M=b\}$ *is* any *partition with*

$$\|\mathcal{Q}\|=\max_{1\leq m\leq M}\triangle y_m<\delta,$$

then

$$U(f;\mathcal{Q}) - L(f;\mathcal{Q}) < \varepsilon.$$

Indeed, since $\triangle y_m < \delta \le \triangle x_n$ for all m and n by choice of δ, each point x_n lies in a *distinct* one of the subintervals $[y_{m-1}, y_m]$ of $\mathcal{Q}$, call it $J_n = [y_{m_n-1}, y_{m_n}]$. The other subintervals $[y_{m-1}, y_m]$ of $\mathcal{Q}$ with m not equal to any of the m_n, each lie in one of the separating intervals $K_n = \left[y_{m_{n-1}}, y_{m_n-1}\right]$ that are formed by the spaces between the intervals J_n. These intervals K_n are the union of one or more consecutive subintervals of $\mathcal{Q}$. We have for each n that

$$\begin{aligned}
&\sum_{m:[y_{m-1},y_m]\subset K_n} \left(\sup_{[y_{m-1},y_m]} f - \inf_{[y_{m-1},y_m]} f \right) \triangle y_m \\
&\le \left(\sup_{\left[y_{m_{n-1}},y_{m_n-1}\right]} f - \inf_{\left[y_{m_{n-1}},y_{m_n-1}\right]} f \right) \sum_{m:[y_{m-1},y_m]\subset K_n} \triangle y_m \\
&\le \left(\sup_{[x_n,x_{n+1}]} f - \inf_{[x_n,x_{n+1}]} f \right) (y_m - y_{m-1}) \\
&\le \left(\sup_{[x_n,x_{n+1}]} f - \inf_{[x_n,x_{n+1}]} f \right) (x_{n+1} - x_n) .
\end{aligned}$$

Summing this in n yields

$$\begin{aligned}
(0.5) \qquad & \sum_{n=1}^{N} \sum_{m:[y_{m-1},y_m]\subset K_n} \left(\sup_{[y_{m-1},y_m]} f - \inf_{[y_{m-1},y_m]} f \right) \triangle y_m \\
\le \ & \sum_{n=1}^{N} \left(\sup_{[x_n,x_{n+1}]} f - \inf_{[x_n,x_{n+1}]} f \right) (x_{n+1} - x_n) = U(f;\mathcal{P}) - L(f;\mathcal{P}) .
\end{aligned}$$

Now we compute

$$\begin{aligned}
U(f;\mathcal{Q}) - L(f;\mathcal{Q}) &= \sum_{m=1}^{M} \left(\sup_{[y_{m-1},y_m]} f - \inf_{[y_{m-1},y_m]} f \right) \triangle y_m \\
&= \sum_{n=1}^{N} \left(\sup_{J_n} f - \inf_{J_n} f \right) (y_{m_n} - y_{m_n-1}) \\
&\quad + \sum_{n=1}^{N} \sum_{m:[y_{m-1},y_m]\subset K_n} \left(\sup_{[y_{m-1},y_m]} f - \inf_{[y_{m-1},y_m]} f \right) \triangle y_m,
\end{aligned}$$

which by (0.5) and choice of δ is dominated by

$$\begin{aligned}
& \mathit{diam}\ f([a,b]) \sum_{n=1}^{N} (y_{m_n} - y_{m_n-1}) + U(f;\mathcal{P}) - L(f;\mathcal{P}) \\
\le \ & \mathit{diam}\ f([a,b])\, N\delta + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

and this proves the claim.

Conversely, if

$$\begin{aligned}
(0.6) \qquad & \text{For every } \varepsilon > 0 \text{ there is } \delta > 0 \text{ such that} \\
& U(f;\mathcal{Q}) - L(f;\mathcal{Q}) < \varepsilon \text{ whenever } \|\mathcal{Q}\| < \delta,
\end{aligned}$$

then the Cauchy criterion (0.4) holds with $\mathcal{P}$ equal to any such $\mathcal{Q}$. Thus (0.6) provides another equivalent definition of the Riemann integral $\int_a^b f$ that is more like the $\varepsilon - \delta$ definition of continuity at a point (compare Definition 4).

1. Properties of the Riemann-Stieltjes integral

The Riemann-Stieltjes integral $\int_a^b f d\alpha$ is a function of the closed interval $[a,b]$, the bounded function f on $[a,b]$, and the nondecreasing function α on $[a,b]$. With respect to each of these three variables, the integral has natural properties related to monotonicity, sums and scalar multiplication. In fact we have the following lemmas dealing with each variable separately, beginning with f, then α and ending with $[a,b]$.

Lemma 7. *Fix $[a,b] \subset \mathbb{R}$ and $\alpha : [a,b] \to \mathbb{R}$ nondecreasing. The set $\mathcal{R}_\alpha [a,b]$ is a real vector space and the integral $\int_a^b f d\alpha$ is a* linear *function of $f \in \mathcal{R}_\alpha [a,b]$: if $f_j \in \mathcal{R}[a,b]$ and $\lambda_j \in \mathbb{R}$, then*

$$f = \lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{R}_\alpha [a,b] \text{ and } \int_a^b f d\alpha = \lambda_1 \int_a^b f_1 d\alpha + \lambda_2 \int_a^b f_2 d\alpha.$$

Furthermore, $\mathcal{R}_\alpha [a,b]$ is partially ordered by declaring $f \leq g$ if $f(x) \leq g(x)$ for $x \in [a,b]$, and the integral $\int_a^b f d\alpha$ is a nondecreasing *function of f with respect to this order: if $f, g \in \mathcal{R}_\alpha [a,b]$ and $f \leq g$, then $\int_a^b f d\alpha \leq \int_a^b g d\alpha$.*

Lemma 8. *Fix $[a,b] \subset \mathbb{R}$ and $f : [a,b] \to \mathbb{R}$ bounded. Then*

$$\mathcal{C}_f [a,b] \equiv \{\alpha : [a,b] \to \mathbb{R} : \alpha \text{ is nondecreasing and } f \in \mathcal{R}_\alpha [a,b]\}$$

is a cone and the integral $\int_a^b f d\alpha$ is a 'positive linear' *function of α: if $\alpha_j \in \mathcal{C}_f [a,b]$ and $c_j \in [0,\infty)$, then*

$$\alpha = c_1 \alpha_1 + c_2 \alpha_2 \in \mathcal{C}_f [a,b] \text{ and } \int_a^b f d\alpha = c_1 \int_a^b f d\alpha_1 + c_2 \int_a^b f d\alpha_2.$$

Lemma 9. *Fix $[a,b] \subset \mathbb{R}$ and $\alpha : [a,b] \to \mathbb{R}$ nondecreasing and $f \in \mathcal{R}_\alpha [a,b]$. If $a < c < b$, then $\alpha : [a,c] \to \mathbb{R}$ and $\alpha : [c,b] \to \mathbb{R}$ are each nondecreasing and*

$$f \in \mathcal{R}_\alpha [a,c] \text{ and } f \in \mathcal{R}_\alpha [c,b] \text{ and } \int_a^b f d\alpha = \int_a^c f d\alpha + \int_c^b f d\alpha.$$

These three lemmas are easy to prove, and are left to the reader. Properties regarding multiplication of functions in $\mathcal{R}_\alpha [a,c]$ and composition of functions are more delicate.

Theorem 19. *Suppose that $f : [a,b] \to [m,M]$ and $f \in \mathcal{R}_\alpha [a,b]$. If $\varphi : [m,M] \to \mathbb{R}$ is continuous, then $\varphi \circ f \in \mathcal{R}_\alpha [a,b]$.*

Corollary 8. *If $f, g \in \mathcal{R}_\alpha [a,b]$, then $fg \in \mathcal{R}_\alpha [a,b]$, $|f| \in \mathcal{R}_\alpha [a,b]$ and*

$$\left| \int_a^b f d\alpha \right| \leq \int_a^b |f| \, d\alpha.$$

Proof: Since $\varphi(x) = x^2$ is continuous, Lemma 7 and Theorem 19 yield

$$fg = \frac{1}{2} \left\{ (f+g)^2 - f^2 - g^2 \right\} \in \mathcal{R}_\alpha [a,b].$$

Since $\varphi(x) = |x|$ is continuous, Theorem 19 yields $|f| \in \mathcal{R}_\alpha[a,b]$. Now choose $c = \pm 1$ so that $c\int_a^b f d\alpha \geq 0$. Then the lemmas imply

$$\left|\int_a^b f d\alpha\right| = c\int_a^b f d\alpha = \int_a^b (cf)\, d\alpha \leq \int_a^b |f|\, d\alpha.$$

Proof (of Theorem 19): Let $h = \varphi \circ f$. We will show that $h \in \mathcal{R}_\alpha[a,b]$ by verifying the Cauchy criterion for integrals (0.4). Fix $\varepsilon > 0$. Since φ is continuous on the compact interval $[m,M]$, it is uniformly continuous on $[m,M]$ by Theorem 14. Thus we can choose $\delta > 0$ such that

$$|\varphi(s) - \varphi(t)| < \varepsilon \text{ whenever } |s-t| < \delta.$$

Since $f \in \mathcal{R}_\alpha[a,b]$, there is by the Cauchy criterion a partition

$$\mathcal{P} = \{a = x_0 < x_1 < \ldots < x_N = b\}$$

such that

$$U(f;\mathcal{P},\alpha) - L(f;\mathcal{P},\alpha) < \delta\varepsilon. \tag{1.1}$$

Let

$$\begin{aligned} M_n &= \sup_{[x_{n-1},x_n]} f \text{ and } m_n = \inf_{[x_{n-1},x_n]} f, \\ M_n^* &= \sup_{[x_{n-1},x_n]} h \text{ and } m_n^* = \inf_{[x_{n-1},x_n]} h, \end{aligned}$$

and set

$$A = \{n : M_n - m_n < \delta\} \text{ and } B = \{n : M_n - m_n \geq \delta\}.$$

The point of the index set A is that for each $n \in A$ we have

$$\begin{aligned} M_n^* - m_n^* &= \sup_{x,y\in[x_{n-1},x_n]} |\varphi(f(x)) - \varphi(f(y))| \leq \sup_{|s-t|\leq M_n - m_n} |\varphi(s) - \varphi(t)| \\ &\leq \sup_{|s-t|<\delta} |\varphi(s) - \varphi(t)| \leq \varepsilon, \qquad n \in A. \end{aligned}$$

As for n in the index set B, we have $\delta \leq M_n - m_n$ and the inequality (1.1) then gives

$$\delta \sum_{n\in B} \triangle\alpha_n \leq \sum_{n\in B} (M_n - m_n) \triangle \alpha_n < \delta\varepsilon.$$

Dividing by $\delta > 0$ we obtain

$$\sum_{n\in B} \triangle\alpha_n < \varepsilon.$$

Now we use the trivial bound

$$M_n^* - m_n^* \leq diam\ \varphi([m,M])$$

to compute that

$$\begin{aligned} U(h;\mathcal{P},\alpha) - L(h;\mathcal{P},\alpha) &= \left\{\sum_{n\in A} + \sum_{n\in B}\right\} (M_n^* - m_n^*) \triangle \alpha_n \\ &\leq \sum_{n\in A} \varepsilon \triangle \alpha_n + \sum_{n\in B} diam\ \varphi([m,M]) \triangle \alpha_n \\ &\leq \varepsilon(\alpha(b) - \alpha(a)) + \varepsilon\ diam\ \varphi([m,M]) \\ &= \varepsilon[\alpha(b) - \alpha(a) + diam\ \varphi([m,M])], \end{aligned}$$

which verifies (0.4) for the existence of $\int_a^b h d\alpha$ as required.

2. The Henstock-Kurtzweil integral

We can reformulate the $\varepsilon - \delta$ definition of the Riemann integral $\int_a^b f$ in (0.6) using a more general notion of partition, that of a tagged partition. If $\mathcal{P} = \{a = x_0 < x_1 < ... < x_N = b\}$ is a partition of $[a,b]$ and we choose points $t_n \in [x_{n-1}, x_n]$ in each subinterval of $\mathcal{P}$, then

$$\mathcal{P}^* = \{a = x_0 \leq t_1 \leq x_1 \leq ... \leq x_{N-1} \leq t_N \leq x_N = b\}, \\ \text{where } x_0 < x_1 < ... < x_N,$$

is called a *tagged* partition $\mathcal{P}^*$ with *underlying* partition $\mathcal{P}$. Thus a tagged partition consists of two finite intertwined sequences $\{x_n\}_{n=0}^N$ and $\{t_n\}_{n=1}^N$, where the sequence $\{x_n\}_{n=0}^N$ is strictly increasing and the sequence $\{t_n\}_{n=1}^N$ need not be. For every tagged partition $\mathcal{P}^*$ of $[a,b]$, define the corresponding *Riemann sum* $S(f;\mathcal{P}^*)$ by

$$S(f;\mathcal{P}^*) = \sum_{n=1}^N f(t_n) \bigtriangleup x_n.$$

Note that $\inf_{[x_{n-1},x_n]} f \leq f(t_n) \leq \sup_{[x_{n-1},x_n]} f$ implies that

$$L(f;\mathcal{P}) \leq S(f;\mathcal{P}^*) \leq U(f;\mathcal{P})$$

for all tagged partitions $\mathcal{P}^*$ with underlying partition $\mathcal{P}$.

Now observe that if $f \in \mathcal{R}[a,b]$, $\varepsilon > 0$ and the partition $\mathcal{P}$ satisfies

$$U(f;\mathcal{P}) - L(f;\mathcal{P}) < \varepsilon,$$

then *every* tagged partition $\mathcal{P}^*$ with underlying partition $\mathcal{P}$ satisfies

$$\left| S(f;\mathcal{P}^*) - \int_a^b f \right| \leq U(f;\mathcal{P}) - L(f;\mathcal{P}) < \varepsilon. \tag{2.1}$$

Conversely if for each $\varepsilon > 0$ there is a partition $\mathcal{P}$ such that *every* tagged partition $\mathcal{P}^*$ with underlying partition $\mathcal{P}$ satisfies (2.1), then (0.4) holds and so $f \in \mathcal{R}[a,b]$.

However, we can also formulate this approach using the $\varepsilon - \delta$ form (0.6) of the definition of $\int_a^b f$. The result is that $f \in \mathcal{R}[a,b]$ if and only if

(2.2) There is $L \in \mathbb{R}$ such that for every $\varepsilon > 0$ there is $\delta > 0$ such that
$$|S(f;\mathcal{P}^*) - L| < \varepsilon \text{ whenever } \|\mathcal{P}^*\| < \delta.$$

Of course if such a number L exists we write $L = \int_a^b f$ and call it the Riemann integral of f on $[a,b]$. Here we define $\|\mathcal{P}^*\|$ to be $\|\mathcal{P}\|$ where $\mathcal{P}$ is the underlying partition of $\mathcal{P}^*$. The reader can easily verify that $f \in \mathcal{R}[a,b]$ if and only if the above condition (2.2) holds.

Now comes the clever insight of Henstock and Kurtzweil. We view the positive constant δ in (2.2) as a *function* on the interval $[a,b]$, and replace it with an arbitrary (not necessarily constant) positive function $\delta : [a,b] \to (0,\infty)$. We refer to such an arbitrary positive function $\delta : [a,b] \to (0,\infty)$ as a *guage* on $[a,b]$. Then for any guage on $[a,b]$, we say that a tagged partition $\mathcal{P}^*$ on $[a,b]$ is δ-*fine* provided

$$[x_{n-1}, x_n] \subset (t_n - \delta(t_n), t_n + \delta(t_n)), \quad 1 \leq n \leq N. \tag{2.3}$$

Thus $\mathcal{P}^*$ is δ-fine if each tag $t_n \in [x_{n-1}, x_n]$ has its associated guage value $\delta(t_n)$ sufficiently large that the open interval centered at t_n with radius $\delta(t_n)$ *contains* the n^{th} subinterval $[x_{n-1}, x_n]$ of the partition $\mathcal{P}$. Now we can give the definition of the Henstock and Kurtzweil integral.

Definition 9. *A function $f : [a,b] \to \mathbb{R}$ is Henstock-Kurtzweil integrable on $[a,b]$, written $f \in \mathcal{HK}[a,b]$, if there is $L \in \mathbb{R}$ such that for every $\varepsilon > 0$ there is a guage $\delta_\varepsilon : [a,b] \to (0,\infty)$ on $[a,b]$ such that*

$$|S(f;\mathcal{P}^*) - L| < \varepsilon \textit{ whenever } \mathcal{P}^* \textit{ is } \delta_\varepsilon\textit{-fine.}$$

It is clear that if $f \in \mathcal{R}[a,b]$ is Riemann integrable, then f satisfies Definition 9 with $L = \int_a^b f$ - simply take δ_ε to be the constant guage δ in (2.2). However, for this new definition to have any value it is necessary that such an L is uniquely determined by Definition 9. This is indeed the case and relies crucially on the fact that $[a,b]$ is compact. Here are the details.

Suppose that Definition 9 holds with both L and L'. Let $\varepsilon > 0$. Then there are guages δ_ε and δ'_ε on $[a,b]$ such that

$$\begin{aligned} |S(f;\mathcal{P}^*) - L| &< \varepsilon \text{ whenever } \mathcal{P}^* \text{ is } \delta_\varepsilon\text{-fine,} \\ |S(f;\mathcal{P}^*) - L'| &< \varepsilon \text{ whenever } \mathcal{P}^* \text{ is } \delta'_\varepsilon\text{-fine.} \end{aligned}$$

Now define

$$\eta_\varepsilon(x) = \min\left\{\delta_\varepsilon(x), \delta'_\varepsilon(x)\right\}, \qquad a \le x \le b.$$

Then η_ε is a guage on $[a,b]$. Here is the critical point: we would like to produce a tagged partition $\mathcal{P}^*_\varepsilon$ that is η_ε-fine! Indeed, if such a tagged partition $\mathcal{P}^*_\varepsilon$ exists, then $\mathcal{P}^*_\varepsilon$ would also be δ_ε-fine and δ'_ε-fine (since $\eta_\varepsilon \le \delta_\varepsilon$ and $\eta_\varepsilon \le \delta'_\varepsilon$) and hence

$$|L - L'| \le |S(f;\mathcal{P}^*_\varepsilon) - L| + |S(f;\mathcal{P}^*_\varepsilon) - L'| < 2\varepsilon$$

for all $\varepsilon > 0$, which forces $L = L'$.

However, if η is any guage on $[a,b]$, let

$$B(x,\eta(x)) = (x - \eta(x), x + \eta(x)) \text{ and } B\left(x, \frac{\eta(x)}{2}\right) = \left(x - \frac{\eta(x)}{2}, x + \frac{\eta(x)}{2}\right).$$

Then $\left\{B\left(x, \frac{\eta(x)}{2}\right)\right\}_{x\in[a,b]}$ is an open cover of the compact set $[a,b]$, hence there is a finite subcover $\left\{B\left(x_n, \frac{\eta(x_n)}{2}\right)\right\}_{n=0}^N$. We may assume that every interval $B\left(x_n, \frac{\eta(x_n)}{2}\right)$ is needed to cover $[a,b]$ by discarding any in turn which are included in the union of the others. We may also assume that $a \le x_0 < x_1 < ... < x_N \le b$. It follows that $B\left(x_{n-1}, \frac{\eta(x_{n-1})}{2}\right) \cap B\left(x_n, \frac{\eta(x_n)}{2}\right) \ne \emptyset$, so the triangle inequality yields

$$|x_n - x_{n-1}| < \frac{\eta(x_{n-1}) + \eta(x_n)}{2}, \qquad 1 \le n \le N.$$

If $\eta(x_n) \ge \eta(x_{n-1})$ then

$$[x_{n-1}, x_n] \subset B(x_n, \eta(x_n)),$$

and so we define

$$t_n = x_n.$$

Otherwise, we have $\eta(x_{n-1}) > \eta(x_n)$ and then

$$[x_{n-1}, x_n] \subset B(x_{n-1}, \eta(x_{n-1})),$$

and so we define

$$t_n = x_{n-1}.$$

The tagged partition

$$\mathcal{P}^* = \{a = x_0 \leq t_1 \leq x_1 \leq ... \leq x_{N-1} \leq t_N \leq x_N = b\}$$

is then η-fine.

With the uniqueness of the Henstock-Kurtzweil integral in hand, and the fact that it extends the definition of the Riemann integral, we can without fear of confusion denote the Henstock-Kurtzweil integral by $\int_a^b f$ when $f \in \mathcal{HK}[a, b]$. It is now possible to develop the standard properties of these integrals as in Theorem 19 and the lemmas above for Riemann integrals. The proofs are typically very similar to those commonly used for Riemann integration. One exception is the Fundamental Theorem of Calculus for the Henstock-Kurtzweil integral, which requires a more complicated proof. In fact, it turns out that the theory of the Henstock-Kurtzweil integral is sufficiently rich to *include* the theory of the Lebesgue integral, which we consider in detail in a later chapter. For further development of the theory of the Henstock-Kurtzweil integral we refer the reader to Bartle and Sherbert [**1**] and the references given there.

CHAPTER 4

Lebesgue measure theory

Recall that f is Riemann integrable on $[0,1)$, written $f \in \mathcal{R}[0,1)$, if $\mathcal{U}(f) = \mathcal{L}(f)$, and we denote the common value by $\int_0^1 f$ or $\int_0^1 f(x)\,dx$. Here $\mathcal{U}(f)$ and $\mathcal{L}(f)$ are the upper and lower Riemann integrals of f on $[0,1)$ respectively given by

$$\begin{aligned} \mathcal{U}(f) &= \inf_{\mathcal{P}} U(f;\mathcal{P}) \equiv \inf_{\mathcal{P}} \sum_{n=1}^{N} \left(\sup_{[x_{n-1},x_n)} f \right) \triangle x_n, \\ \mathcal{L}(f) &= \sup_{\mathcal{P}} L(f;\mathcal{P}) \equiv \sup_{\mathcal{P}} \sum_{n=1}^{N} \left(\inf_{[x_{n-1},x_n)} f \right) \triangle x_n, \end{aligned}$$

where $\mathcal{P} = \{0 = x_0 < x_1 < ... < x_N = 1\}$ is any partition of $[0,1)$ and $\triangle x_n = x_n - x_{n-1} > 0$. For convenience we work with $[0,1)$ in place of $[0,1]$ for now.

This definition is simple and easy to work with and applies in particular to bounded continuous functions f on $[0,1)$ since it is not too hard to prove that $f \in \mathcal{R}[0,1)$ for such f. However, if we consider the vector space $L^2_{\mathcal{R}}([0,1))$ of Riemann integrable functions $f \in \mathcal{R}[0,1)$ endowed with the metric

$$d(f,g) = \left(\int_0^1 |f(x) - g(x)|^2 \, dx \right)^{\frac{1}{2}},$$

it turns out that while $L^2_{\mathcal{R}}([0,1))$ can indeed be proved a metric space (actually we must consider equivalence classes of functions where we identify functions f and g if $\int_0^1 |f(x) - g(x)|^2 \, dx = 0$), it fails to be *complete*. This is a serious shortfall of Riemann's theory of integration, and is our main motivation for considering the more complicated theory of Lebesgue below. We note that the immediate reason for the lack of completeness of $L^2_{\mathcal{R}}([0,1))$ is the inability of Riemann's theory to handle general unbounded functions. For example, the sequence $\{f_n\}_{n=1}^{\infty}$ defined on $[0,1)$ by

$$f_n(x) = \chi_{[0,1)}(x) \min\left\{ x^{-\frac{1}{4}}, 2^{\frac{n}{4}} \right\}$$

is a Cauchy sequence in $L^2_{\mathcal{R}}([0,1))$ that clearly has no bounded function as limit in $L^2_{\mathcal{R}}([0,1))$. Indeed,

$$d(f_n, f_{n+1})^2 = \int_0^{2^{-n}} |f_{n+1}(x) - f_n(x)|^2 \, dx \le \int_0^{2^{-n}} \left| 2^{\frac{n}{4}} \right|^2 dx = 2^{-\frac{n}{2}}$$

and so for $m > n$ we have

$$d(f_n, f_m) \le \sum_{k=n}^{m-1} d(f_k, f_{k+1}) \le \sum_{k=n}^{m-1} 2^{-\frac{k}{4}} \to 0 \text{ as } n \to \infty.$$

However, even locally there are problems. For example, once we have Lebesgue's theory in hand, we can construct a famous example of a Lebesgue measurable subset E of $[0,1)$ with the (somewhat surprising) property that

$$0 < |E \cap (a,b)| < b-a, \qquad 0 \leq a < b \leq 1,$$

where $|F|$ denotes the Lebesgue measure of a measurable set F (see Problem 3 below). It follows that the characteristic function χ_E is bounded and Lebesgue measurable, but that there is no Riemann integrable function f such that $f = \chi_E$ almost everywhere, since such an f would satisfy $\mathcal{U}(f) = 1$ and $\mathcal{L}(f) = 0$. Nevertheless, by Lusin's Theorem (see page 34 in [**6**] or page 55 in [**4**]) there is a sequence of compactly supported *continuous* functions (hence Riemann integrable) converging to χ_E almost everywhere and that are uniformly bounded. By the Dominated Convergence Theorem below, this sequence is Cauchy in $L^2_{\mathcal{R}}([0,1))$.

On the other hand, in *Lebesgue's* theory of integration, we partition the *range* $[0,M)$ of the bounded function f into a homogeneous partition,

$$[0,M) = \bigcup_{n=1}^{N} \left[(n-1)\frac{M}{N}, n\frac{M}{N}\right) \equiv \bigcup_{n=1}^{N} I_n,$$

and we consider the associated upper and lower *Lebesgue* sums of f on $[0,1)$ defined by

$$\begin{aligned} U^*(f;\mathcal{P}) &= \sum_{n=1}^{N}\left(n\frac{M}{N}\right)\left|f^{-1}(I_n)\right|, \\ L^*(f;\mathcal{P}) &= \sum_{n=1}^{N}\left((n-1)\frac{M}{N}\right)\left|f^{-1}(I_n)\right|, \end{aligned}$$

where of course

$$f^{-1}(I_n) = \left\{x \in [0,1) : f(x) \in I_n = \left[(n-1)\frac{M}{N}, n\frac{M}{N}\right)\right\},$$

and $|E|$ denotes the "measure" or "length" of the subset E of $[0,1)$.

Here there will be no problem obtaining that $U^*(f;\mathcal{P}) - L^*(f;\mathcal{P})$ is small provided we can make sense of $\left|f^{-1}(I_n)\right|$. But this is precisely the difficulty with Lebesgue's approach - we need to define a notion of "measure" or "length" for subsets E of $[0,1)$. That this is not going to be as easy as we might hope is evidenced by the following negative result. Let $\mathcal{P}([0,1))$ denote the power set of $[0,1)$, i.e. the set of all subsets of $[0,1)$. For $x \in [0,1)$ and $E \in \mathcal{P}([0,1))$ we define the translation $E \oplus x$ of E by x to be the set in $\mathcal{P}([0,1))$ defined by

$$\begin{aligned} E \oplus x &= E + x \pmod 1 \\ &= \{z \in [0,1) : \text{there is } y \in E \text{ with } y + x - z \in \mathbb{Z}\}. \end{aligned}$$

Theorem 20. *There is* no *map* $\mu : \mathcal{P}([0,1)) \to [0,\infty)$ *satisfying the following three properties:*

(1) $\mu([0,1)) = 1$,
(2) $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$ *whenever* $\{E_n\}_{n=1}^{\infty}$ *is a pairwise disjoint sequence of sets in* $\mathcal{P}([0,1))$,
(3) $\mu(E \oplus x) = \mu(E)$ *for all* $E \in \mathcal{P}([0,1))$.

Remark 7. *All three of these properties are desirable for any notion of measure or length of subsets of* $[0,1)$. *The theorem suggests then that we should* not *demand that every subset of* $[0,1)$ *be "measurable". This will then restrict the functions* f *that we can integrate to those for which* $f^{-1}([a,b))$ *is "measurable" for all* $-\infty < a < b < \infty$.

Proof: Let $\{r_n\}_{n=1}^{\infty} = \mathbb{Q} \cap [0,1)$ be an enumeration of the rational numbers in $[0,1)$. Define an equivalence relation on $[0,1)$ by declaring that $x \sim y$ if $x - y \in \mathbb{Q}$. Let $\mathcal{A}$ be the set of equivalence classes. Use the *axiom of choice* to pick a representative $a = \langle A \rangle$ from each equivalence class A in $\mathcal{A}$. Finally, let $E = \{\langle A \rangle : A \in \mathcal{A}\}$ be the set consisting of these representatives a, one from each equivalence class A in $\mathcal{A}$.

Then we have

$$[0,1) = \dot{\bigcup}_{n=1}^{\infty} E \oplus r_n.$$

Indeed, if $x \in [0,1)$, then $x \in A$ for some $A \in \mathcal{A}$, and thus $x \sim a = \langle A \rangle$, i.e. $x - a \in \{r_n\}_{n=1}^{\infty}$. If $x \geq a$ then $x - a \in \mathbb{Q} \cap [0,1)$ and $x = a + r_m$ where $a \in E$ and $r_m \in \{r_n\}_{n=1}^{\infty}$. If $x < a$ then $x - a + 1 \in \mathbb{Q} \cap [0,1)$ and $x = a + (r_m \ominus 1)$ where $a \in E$ and $r_m \ominus 1 \in \{r_n\}_{n=1}^{\infty}$. Finally, if $a \oplus r_m = b \oplus r_n$, then $a \ominus b = r_n \ominus r_m \in \mathbb{Q}$ which implies that $a \sim b$ and then $r_n = r_m$.

Now by properties (1), (2) and (3) in succession we have

$$1 = \mu([0,1)) = \mu\left(\dot{\bigcup}_{n=1}^{\infty} E \oplus r_n\right) = \sum_{n=1}^{\infty} \mu(E \oplus r_n) = \sum_{n=1}^{\infty} \mu(E),$$

which is impossible since the infinite series $\sum_{n=1}^{\infty} \mu(E)$ is either ∞ if $\mu(E) > 0$ or 0 if $\mu(E) = 0$.

1. Lebesgue measure on the real line

In order to define a "measure" satisfying the three properties in Theorem 20, we must restrict the domain of definition of the set functional μ to a "suitable" proper subset of the power set $\mathcal{P}([0,1))$. A good notion of "suitable" is captured by the following definition where we expand our quest for measure to the entire real line.

Definition 10. *A collection* $\mathcal{A} \subset \mathcal{P}(\mathbb{R})$ *of subsets of real numbers* $\mathbb{R}$ *is called a* σ-algebra *if the following properties are satisfied:*

1. $\phi \in \mathcal{A}$,
2. $A^c \in \mathcal{A}$ *whenever* $A \in \mathcal{A}$,
3. $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ *whenever* $A_n \in \mathcal{A}$ *for all* n.

Here is the theorem asserting the existence of "Lebesgue measure" on the real line.

Theorem 21. *There is a* σ*-algebra* $\mathcal{L} \subset \mathcal{P}(\mathbb{R})$ *and a function* $\mu : \mathcal{L} \to [0,\infty]$ *such that*

1. $[a,b) \in \mathcal{L}$ *and* $\mu([a,b)) = b - a$ *for all* $-\infty < a < b < \infty$,
2. $\dot{\bigcup}_{n=1}^{\infty} E_n \in \mathcal{L}$ *and* $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$ *whenever* $\{E_n\}_{n=1}^{\infty}$ *is a pairwise disjoint sequence of sets in* $\mathcal{L}$,
3. $E + x \in \mathcal{L}$ *and* $\mu(E + x) = \mu(E)$ *for all* $E \in \mathcal{L}$,

(4) $E \in \mathcal{L}$ *and* $\mu(E) = 0$ *whenever* $E \subset F$ *and* $F \in \mathcal{L}$ *with* $\mu(F) = 0$.

The sets in the σ-algebra $\mathcal{L}$ are called *Lebesgue measurable* sets. A pair $(\mathcal{L}, \mu)$ satisfying only property (2) is called a measure space. Property (1) says that the measure μ is an extension of the usual length function on intervals. Property (3) says that the measure is translation invariant, while property (4) says that the measure is complete.

From property (2) and the fact that μ is nonnegative, and finite on intervals, we easily obtain the following elementary consequences (where membership in $\mathcal{L}$ is implied by *context*):

$$\begin{aligned} \phi &\in \mathcal{L} \text{ and } \mu(\phi) = 0, \\ E &\in \mathcal{L} \text{ for every open set } E \text{ in } \mathbb{R}, \\ \mu(I) &= b - a \text{ for } \textit{any} \text{ interval } I \text{ with endpoints } a \text{ and } b, \\ \mu(E) &= \sup_n \mu(E_n) = \lim_{n\to\infty} \mu(E_n) \text{ if } E_n \nearrow E, \\ \mu(E) &= \inf_n \mu(E_n) = \lim_{n\to\infty} \mu(E_n) \text{ if } E_n \searrow E \textbf{ and } \mu(E_1) < \infty. \end{aligned} \tag{1.1}$$

For example, the fourth line follows from writing

$$E = E_1 \dot{\cup} \left\{ \dot{\bigcup}_{n=1}^{\infty} E_{n+1} \cap (E_n)^c \right\}$$

and then using property (2) of μ.

To prove Theorem 21 we follow the treatment in [**6**] with simplifications due to the fact that Theorem 15 implies the connected open subsets of the real numbers $\mathbb{R}$ are just the open intervals (a, b). Define for *any* $E \in \mathcal{P}(\mathbb{R})$, the *outer Lebesgue measure* $\mu^*(E)$ of E by,

$$\mu^*(E) = \inf \left\{ \sum_{n=1}^{\infty} (b_n - a_n) : E \subset \dot{\bigcup}_{n=1}^{\infty} (a_n, b_n) \text{ and } -\infty \le a_n < b_n \le \infty \right\}. \tag{1.2}$$

It is immediate that μ^* is monotone,

$$\mu^*(E) \le \mu^*(F) \text{ if } E \subset F.$$

A little less obvious is countable subadditivity of μ^*. The reason lies in the use of *pairwise disjoint* covers of E by open intervals in the definition of $\mu^*(E)$ in (1.2). If we had instead used arbitrary open covers by open intervals in the definition, then countable subadditivity of μ^* would have been trivial.

LEMMA 10. μ^* *is countably subadditive:*

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n \right) \le \sum_{n=1}^{\infty} \mu^*(E_n), \qquad \{E_n\}_{n=1}^{\infty} \subset \mathcal{P}(\mathbb{R}).$$

Proof: Given $0 < \varepsilon < 1$, we have $E_n \subset \dot{\bigcup}_{k=1}^{\infty} (a_{k,n}, b_{k,n})$ with

$$\sum_{k=1}^{\infty} (b_{k,n} - a_{k,n}) < \mu^*(E_n) + \frac{\varepsilon}{2^n}, \qquad n \ge 1.$$

Now let

$$\bigcup_{n=1}^{\infty}\left(\dot{\bigcup}_{k=1}^{\infty}\left(a_{k,n},b_{k,n}\right)\right)=\dot{\bigcup}_{m=1}^{M^*}\left(c_m,d_m\right),$$

where $M^* \in \mathbb{N}\cup\{\infty\}$. Then define disjoint sets of indices

$$\mathcal{I}_m=\left\{(k,n):\left(a_{k,n},b_{k,n}\right)\subset\left(c_m,d_m\right)\right\}.$$

In the case $c_m, d_m \in \mathbb{R}$, we can choose by compactness a finite subset $\mathcal{F}_m$ of $\mathcal{I}_m$ such that

$$\left[c_m+\frac{\varepsilon}{2}\delta_m,d_m-\frac{\varepsilon}{2}\delta_m\right]\subset\bigcup_{(k,n)\in\mathcal{F}_m}^{\infty}\left(a_{k,n},b_{k,n}\right), \tag{1.3}$$

where $\delta_m = d_m - c_m$. We may assume that each such interval $(a_{k,n}, b_{k,n})$ has nonempty intersection with the compact interval on the left side of (1.3). Fix m and arrange the left endpoints $\{a_{k,n}\}_{(k,n)\in\mathcal{F}_m}$ in strictly increasing order $\{a_i\}_{i=1}^{I}$ and denote the corresponding right endpoints by b_i (if there is more than one interval (a_i, b_i) with the same left endpoint a_i, discard all but one of the largest of them). From (1.3) it now follows that $a_{i+1} \in (a_i, b_i)$ for $i < I$ since otherwise b_i would be in the left side of (1.3), but not in the right side, a contradiction. Thus $a_{i+1} - a_i \leq b_i - a_i$ for $1 \leq i < I$ and we have the inequality

$$\begin{aligned}(1-\varepsilon)\,\delta_m &= \left(d_m-\frac{\varepsilon}{2}\delta_m\right)-\left(c_m+\frac{\varepsilon}{2}\delta_m\right)\\ &\leq b_I-a_1=\left(b_I-a_I\right)+\sum_{i=1}^{I-1}\left(a_{i+1}-a_i\right)\\ &\leq \sum_{i=1}^{I}\left(b_i-a_i\right)\leq\sum_{(k,n)\in\mathcal{F}_m}\left(b_{k,n}-a_{k,n}\right)\\ &\leq \sum_{(k,n)\in\mathcal{I}_m}\left(b_{k,n}-a_{k,n}\right).\end{aligned}$$

We also observe that a similar argument shows that $\sum_{(k,n)\in\mathcal{I}_m}(b_{k,n}-a_{k,n})=\infty$ if $\delta_m = \infty$. Then we have

$$\begin{aligned}\mu^*(E) &\leq \sum_{m=1}^{\infty}\delta_m\leq\frac{1}{1-\varepsilon}\sum_{m=1}^{\infty}\sum_{(k,n)\in\mathcal{F}_m}\left(b_{k,n}-a_{k,n}\right)\\ &\leq \frac{1}{1-\varepsilon}\sum_{k,n}\left(b_{k,n}-a_{k,n}\right)=\frac{1}{1-\varepsilon}\sum_{n=1}^{\infty}\sum_{k=1}^{\infty}\left(b_{k,n}-a_{k,n}\right)\\ &< \frac{1}{1-\varepsilon}\sum_{n=1}^{\infty}\left(\mu^*\left(E_n\right)+\frac{\varepsilon}{2^n}\right)=\frac{1}{1-\varepsilon}\sum_{n=1}^{\infty}\mu^*\left(E_n\right)+\frac{\varepsilon}{1-\varepsilon}.\end{aligned}$$

Let $\varepsilon \to 0$ to obtain the countable subadditivity of μ^*.

Definition 11. *Now define the subset $\mathcal{L}$ of $\mathcal{P}(\mathbb{R})$ to consist of all subsets A of the real line such that for every $\varepsilon > 0$, there is an open set $G \supset A$ satisfying*

$$\mu^*\left(G\setminus A\right)<\varepsilon. \tag{1.4}$$

Remark 8. *Condition (1.4) says that A can be well approximated from the outside by open sets. The most difficult task we will face below in using this definition of $\mathcal{L}$ is to prove that such sets A can also be well approximated from the inside by closed sets.*

Set

$$\mu(A) = \mu^*(A), \qquad A \in \mathcal{L}.$$

Trivially, every open set and every interval is in $\mathcal{L}$. We will use the following two claims in the proof of Theorem 21.

Claim 2. *If G is open and $G = \dot{\bigcup}_{n=1}^{N^*}(a_n, b_n)$ (where $N^* \in \mathbb{N} \cup \{\infty\}$) is the decomposition of G into its connected components (a_n, b_n) (Proposition 2 of Chapter 2), then*

$$\mu(G) = \mu^*(G) = \sum_{n=1}^{N^*}(b_n - a_n).$$

We first prove Claim 2 when $N^* < \infty$. If $G \subset \dot{\bigcup}_{m=1}^{\infty}(c_m, d_m)$, then for each $1 \leq n \leq N^*$, $(a_n, b_n) \subset (c_m, d_m)$ for some m since (a_n, b_n) is connected. If

$$\mathcal{I}_m = \{n : (a_n, b_n) \subset (c_m, d_m)\},$$

it follows upon arranging the a_n in increasing order that

$$\sum_{n \in \mathcal{I}_m}(b_n - a_n) \leq d_m - c_m,$$

since the intervals (a_n, b_n) are pairwise disjoint. We now conclude that

$$\begin{aligned} \mu^*(G) &= \inf\left\{\sum_{m=1}^{\infty}(d_m - c_m) : G \subset \dot{\bigcup}_{m=1}^{\infty}(c_m, d_m)\right\} \\ &\geq \sum_{m=1}^{\infty}\sum_{n \in \mathcal{I}_m}(b_n - a_n) = \sum_{n=1}^{N^*}(b_n - a_n), \end{aligned}$$

and hence that $\mu^*(G) = \sum_{n=1}^{N^*}(b_n - a_n)$ by definition since $G \subset \dot{\bigcup}_{m=1}^{N^*}(a_n, b_n)$.

Finally, if $N^* = \infty$, then from what we just proved and monotonicity, we have

$$\mu^*(G) \geq \mu^*\left(\dot{\bigcup}_{m=1}^{N}(a_n, b_n)\right) = \sum_{n=1}^{N}(b_n - a_n)$$

for each $1 \leq N < \infty$. Taking the supremum over N gives $\mu^*(G) \geq \sum_{n=1}^{\infty}(b_n - a_n)$, and then equality follows by definition since $G \subset \dot{\bigcup}_{n=1}^{\infty}(a_n, b_n)$.

Claim 3. *If A and B are disjoint compact subsets of $\mathbb{R}$, then*

$$\mu^*(A) + \mu^*(B) = \mu^*(A \cup B).$$

First note that

$$\delta = dist(A, B) \equiv \inf\{|x - y| : x \in A, y \in B\} > 0,$$

since the function $f(x, y) \equiv |x - y|$ is positive and continuous on the closed and bounded (hence compact) subset $A \times B$ of the plane - Theorem 12 shows that f

achieves its infimum *dist* (A,B), which is thus positive. So we can find open sets U and V such that

$$A \subset U \text{ and } B \subset V \text{ and } U \cap V = \phi.$$

For example, $U = \bigcup_{x \in A} B\left(x, \frac{\delta}{2}\right)$ and $V = \bigcup_{x \in B} B\left(x, \frac{\delta}{2}\right)$ work. Now suppose that

$$A \cup B \subset G \equiv \dot{\bigcup}_{n=1}^{\infty} (a_n, b_n).$$

Then we have

$$A \subset U \cap G = \dot{\bigcup}_{k=1}^{K^*} (e_k, f_k) \text{ and } B \subset V \cap G = \dot{\bigcup}_{\ell=1}^{L^*} (g_\ell, h_\ell),$$

and then from Claim 2 and monotonicity of μ^* we obtain, using that G is a *disjoint* union of $G \cap U$ and $G \cap V$,

$$\begin{aligned} \mu^*(A) + \mu^*(B) &\leq \sum_{k=1}^{K^*} (f_k - e_k) + \sum_{\ell=1}^{L^*} (h_\ell - g_\ell) \\ &= \mu^*\left(\left(\dot{\bigcup}_{k=1}^{K^*} (e_k, f_k)\right) \dot{\cup} \left(\dot{\bigcup}_{\ell=1}^{L^*} (g_\ell, h_\ell)\right)\right) \\ &\leq \mu^*(G) = \sum_{n=1}^{\infty} (b_n - a_n). \end{aligned}$$

Taking the infimum over such G gives $\mu^*(A) + \mu^*(B) \leq \mu^*(A \cup B)$, and subadditivity of μ^* now proves equality.

Proof (of Theorem 21): We now prove that $\mathcal{L}$ is a σ-algebra and that $\mathcal{L}$ and μ satisfy the four properties in the statement of Theorem 21. First we establish that $\mathcal{L}$ is a σ-algebra in four steps.

Step 1: $A \in \mathcal{L}$ if $\mu^*(A) = 0$.

Given $\varepsilon > 0$, there is an open $G \supset A$ with $\mu^*(G) < \varepsilon$. But then $\mu^*(G \setminus A) \leq \mu^*(G) < \varepsilon$ by monontonicity.

Step 2: $\bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$ whenever $A_n \in \mathcal{L}$ for all n.

Given $\varepsilon > 0$, there is an open $G_n \supset A_n$ with $\mu^*(G_n \setminus A_n) < \frac{\varepsilon}{2^n}$. Then $A \equiv \bigcup_{n=1}^{\infty} A_n$ is contained in the open set $G \equiv \bigcup_{n=1}^{\infty} G_n$, and since $G \setminus A$ is contained in $\bigcup_{n=1}^{\infty} (G_n \setminus A_n)$, monotonicity and subadditivity of μ^* yield

$$\mu^*(G \setminus A) \leq \mu^*\left(\bigcup_{n=1}^{\infty} (G_n \setminus A_n)\right) \leq \sum_{n=1}^{\infty} \mu^*(G_n \setminus A_n) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon.$$

Step 3: $A \in \mathcal{L}$ if A is closed.

Suppose first that A is compact, and let $\varepsilon > 0$. Then using Claim 2 there is $G = \dot{\bigcup}_{n=1}^{N^*} (a_n, b_n)$ containing A with

$$\mu^*(G) = \sum_{n=1}^{\infty} (b_n - a_n) \leq \mu^*(A) + \varepsilon < \infty.$$

Now $G \setminus A$ is open and so $G \setminus A = \dot{\bigcup}_{m=1}^{M^*} (c_m, d_m)$ by Proposition 2. We want to show that $\mu^* (G \setminus A) \leq \varepsilon$. Fix a finite $M \leq M^*$ and

$$0 < \eta < \frac{1}{2} \min_{1 \leq m \leq M} (d_m - c_m) .$$

Then the compact set

$$K_\eta = \bigcup_{m=1}^{M} [c_m + \eta, d_m - \eta]$$

is disjoint from A, so by Claim 3 and induction we have

$$\mu^* (A \cup K_\eta) = \mu^* (A) + \mu^* (K_\eta) = \mu^* (A) + \sum_{m=1}^{M} \mu^* ([c_m + \eta, d_m - \eta]) .$$

We conclude from subadditivity and $A \cup K_\eta \subset G$ that

$$\mu^* (A) + \sum_{m=1}^{M} (d_m - c_m - 2\eta) = \mu^* (A \cup K_\eta) \leq \mu^* (G) \leq \mu^* (A) + \varepsilon.$$

Since $\mu^* (A) < \infty$ for A compact, we thus have

$$\sum_{m=1}^{M} (d_m - c_m) \leq \varepsilon + 2M\eta$$

for all $0 < \eta < \frac{1}{2} \min_{1 \leq m \leq M} (d_m - c_m)$. Hence $\sum_{m=1}^{M} (d_m - c_m) \leq \varepsilon$ and taking the supremum in $M \leq M^*$ we obtain from Claim 2 that

$$\mu^* (G \setminus A) = \sum_{m=1}^{M^*} (d_m - c_m) \leq \varepsilon.$$

Finally, if A is closed, it is a countable union of compact sets $A = \bigcup_{n=1}^{\infty} ([-n, n] \cap A)$, and hence $A \in \mathcal{L}$ by Step 2.

Step 4: $A^c \in \mathcal{L}$ if $A \in \mathcal{L}$.

For each $n \geq 1$ there is by Claim 2 an open set $G_n \supset A$ such that $\mu^* (G_n \setminus A) < \frac{1}{n}$. Then $F_n \equiv G_n^c$ is closed and hence $F_n \in \mathcal{L}$ by Step 3. Thus

$$S \equiv \bigcup_{n=1}^{\infty} F_n \in \mathcal{L}, \quad S \subset A^c,$$

and $A^c \setminus S \subset G_n \setminus A$ for all n implies that

$$\mu^* (A^c \setminus S) \leq \mu^* (G_n \setminus A) < \frac{1}{n}, \qquad n \geq 1.$$

Thus $\mu^* (A^c \setminus S) = 0$ and by Step 1 we have $A^c \setminus S \in \mathcal{L}$. Finally, Step 2 shows that

$$A^c = S \cup (A^c \setminus S) \in \mathcal{L}.$$

Thus far we have shown that $\mathcal{L}$ is a σ-algebra, and we now turn to proving that $\mathcal{L}$ and μ satisfy the four properties in Theorem 21. Property (1) is an easy exercise. Property (2) is the main event. Let $\{E_n\}_{n=1}^{\infty}$ be a pairwise disjoint sequence of sets in $\mathcal{L}$, and let $E = \dot{\bigcup}_{n=1}^{\infty} E_n$.

We will consider first the case where each of the sets E_n is bounded. Let $\varepsilon > 0$ be given. Then $E_n^c \in \mathcal{L}$ and so there are open sets $G_n \supset E_n^c$ such that

$$\mu^*\left(G_n \setminus E_n^c\right) < \frac{\varepsilon}{2^n}, \qquad n \geq 1.$$

Equivalently, with $F_n = G_n^c$, we have F_n closed, contained in E_n, and

$$\mu^*\left(E_n \setminus F_n\right) < \frac{\varepsilon}{2^n}, \qquad n \geq 1.$$

Thus the sets F_n in the sequence $\{F_n\}_{n=1}^\infty$ are compact and pairwise disjoint. Claim 3 and induction shows that

$$\sum_{n=1}^N \mu^*\left(F_n\right) = \mu^*\left(\bigcup_{n=1}^N F_n\right) \leq \mu^*\left(E\right), \qquad N \geq 1,$$

and taking the supremum over N yields

$$\sum_{n=1}^\infty \mu^*\left(F_n\right) \leq \mu^*\left(E\right).$$

Thus we have

$$\begin{aligned} \sum_{n=1}^\infty \mu^*\left(E_n\right) &\leq \sum_{n=1}^\infty \left\{\mu^*\left(E_n \setminus F_n\right) + \mu^*\left(F_n\right)\right\} \\ &\leq \sum_{n=1}^\infty \frac{\varepsilon}{2^n} + \sum_{n=1}^\infty \mu^*\left(F_n\right) \leq \varepsilon + \mu^*\left(E\right). \end{aligned}$$

Since $\varepsilon > 0$ we conclude that $\sum_{n=1}^\infty \mu^*\left(E_n\right) \leq \mu^*\left(E\right)$, and subadditivity of μ^* then proves equality.

In general, define $E_{n,k} = E_n \cap [k, k+1)$ for $k \in \mathbb{Z}$ so that

$$E = \dot{\bigcup}_{n=1}^\infty E_n = \dot{\bigcup}_{n\geq 1, k\in\mathbb{Z}} E_{n,k}.$$

Then from what we just proved applied first to E and then to E_n we have

$$\mu^*\left(E\right) = \sum_{n\geq 1, k\in\mathbb{Z}} \mu^*\left(E_{n,k}\right) = \sum_{n=1}^\infty \left(\sum_{k\in\mathbb{Z}} \mu^*\left(E_{n,k}\right)\right) = \sum_{n=1}^\infty \mu^*\left(E_n\right).$$

Finally, property (3) follows from the observation that $E \subset \dot{\bigcup}_{n=1}^\infty (a_n, b_n)$ if and only if $E + x \subset \dot{\bigcup}_{n=1}^\infty (a_n + x, b_n + x)$. It is then obvious that $\mu^*\left(E + x\right) = \mu^*\left(E\right)$ and that $E + x \in \mathcal{L}$ if $E \in \mathcal{L}$. Property (4) is immediate from Step 1 above. This completes the proof of Theorem 21.

REMARK 9. *The above proof also establishes the* regularity *of Lebesgue measure: for every $E \in \mathcal{L}$ and $\varepsilon > 0$, there is a closed set F and an open set G satisfying*

$$\begin{aligned} F &\subset E \subset G, \\ \mu\left(G \setminus F\right) &< \varepsilon. \end{aligned}$$

This follows from the definition of $\mathcal{L}$ together with the fact that $\mathcal{L}$ is closed under complementation.

EXERCISE 1. *Use the regularity of Lebesgue measure to show that* $E \in \mathcal{L}$ if and only if *there is an increasing sequence* $\{K_n\}_{n=1}^{\infty}$ *of compact sets in* $\mathbb{R}$ *and a null set* N *(i.e.* $\mu^*(N) = 0$*) such that*

$$E = \left(\bigcup_{n=1}^{\infty} K_n \right) \cup N.$$

Show also that if another pair $(\mathcal{L}', \mu')$ *satisfies (1) - (4), then* $K \in \mathcal{L}'$ *and* $\mu'(K) = \mu(K)$ *for all compact subsets* K *of* $\mathbb{R}$*. Deduce from this that* $(\mathcal{L}', \mu')$ *is an extension of* $(\mathcal{L}, \mu)$*, i.e.* $\mathcal{L}' \supset \mathcal{L}$ *and* $\mu'(E) = \mu(E)$ *for all* $E \in \mathcal{L}$*.*

2. Measurable functions and integration

Let $[-\infty, \infty] = \mathbb{R} \cup \{-\infty, \infty\}$ be the extended real numbers with order and (some) algebra operations defined by

$$\begin{aligned} -\infty &< x < \infty, \quad x \in \mathbb{R}, \\ x + \infty &= \infty, \quad x \in \mathbb{R}, \\ x - \infty &= -\infty, \quad x \in \mathbb{R}, \\ x \cdot \infty &= \infty, \quad x > 0, \\ x \cdot \infty &= -\infty, \quad x < 0, \\ 0 \cdot \infty &= 0. \end{aligned}$$

The final assertion $0 \cdot \infty = 0$ is dictated by $\sum_{n=1}^{\infty} a_n = 0$ if all the $a_n = 0$. It turns out that these definitions give rise to a consistent theory of measure and integration of functions with values in the extended real number system.

Let $f : \mathbb{R} \to [-\infty, \infty]$. We say that f is (Lebesgue) measurable if

$$f^{-1}([-\infty, x)) \in \mathcal{L}, \quad x \in \mathbb{R}.$$

The simplest examples of measurable functions are the characteristic functions χ_E of measurable sets E. Indeed,

$$(\chi_E)^{-1}([-\infty, x)) = \begin{cases} \phi & \text{if} \quad x \le 0 \\ E^c & \text{if} \quad 0 < x \le 1 \\ \mathbb{R} & \text{if} \quad x > 1 \end{cases}.$$

It is then easy to see that finite linear combinations $s = \sum_{n=1}^{N} a_n \chi_{E_n}$ of such characteristic functions χ_{E_n}, called simple functions, are also measurable. Here $a_n \in \mathbb{R}$ and E_n is a measurable subset of $\mathbb{R}$. Note that these functions are those arising as upper and lower Lebesgue sums. However, since the difference of upper and lower Lebesgue sums is automatically controlled, we proceed to develop integration by an approximation method instead. It turns out that if we define the integral of a simple function $s = \sum_{n=1}^{N} a_n \chi_{E_n}$ by

$$\int_{\mathbb{R}} s = \sum_{n=1}^{N} a_n \mu(E_n),$$

the value is independent of the representation of s as a simple function. Armed with this fact we can then extend the definition of integral $\int_{\mathbb{R}} f$ to functions f that are nonnegative on $\mathbb{R}$, and then to functions f such that $\int_{\mathbb{R}} |f| < \infty$.

At each stage one establishes the relevant properties of the integral along with the most useful theorems. For the most part these extensions are rather routine, the cleverness inherent in the theory being in the overarching organization of the concepts rather than in the details of the demonstrations. As a result, we will merely state the main results in logical order and sketch proofs when not simply routine. We will however give fairly detailed proofs of the three famous convergence theorems, the Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem. The reader is referred to the excellent exposition in [**6**] for the complete story including many additional fascinating insights.

2.1. Properties of measurable functions. From now on we denote the Lebesgue measure of a measurable subset E of $\mathbb{R}$ by $|E|$ rather than by $\mu(E)$ as in the previous sections. We say that two measurable functions $f, g : \mathbb{R} \to [-\infty, \infty]$ are equal *almost everywhere* (often abbreviated *a.e.*) if

$$|\{x \in \mathbb{R} : f(x) \neq g(x)\}| = 0.$$

We say that f is *finite-valued* if $f : \mathbb{R} \to \mathbb{R}$. We now collect a number of elementary properties of measurable functions.

Lemma 11. *Suppose that $f, f_n, g : \mathbb{R} \to [-\infty, \infty]$ for $n \in \mathbb{N}$.*

(1) *If f is finite-valued, then f is measurable* if and only if *$f^{-1}(G) \in \mathcal{L}$ for all open sets $G \subset \mathbb{R}$* if and only if *$f^{-1}(F) \in \mathcal{L}$ for all closed sets $F \subset \mathbb{R}$.*
(2) *If f is finite-valued and continuous, then f is measurable.*
(3) *If f is finite-valued and measurable and $\Phi : \mathbb{R} \to \mathbb{R}$ is continuous, then $\Phi \circ f$ is measurable.*
(4) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions, then the following functions are all measurable:*

$$\sup_n f_n(x), \quad \inf_n f_n(x), ... \limsup_{n\to\infty} f_n(x), \quad \liminf_{n\to\infty} f_n(x).$$

(5) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions and $f(x) = \lim_{n\to\infty} f_n(x)$, then f is measurable.*
(6) *If f is measurable, so is f^n for $n \in \mathbb{N}$.*
(7) *If f and g are finite-valued and measurable, then so are $f + g$ and fg.*
(8) *If f is measurable and $f = g$ almost everywhere, then g is measurable.*

Comments: For property (1), first show that f is measurable if and only if $f^{-1}((a,b)) \in \mathcal{L}$ for all $-\infty < a < b < \infty$. For property (3) use $(\Phi \circ f)^{-1}(G) = f^{-1}\left(\Phi^{-1}(G)\right)$ and note that $\Phi^{-1}(G)$ is open if G is open. For property (7), use

$$\begin{aligned} \{f + g > a\} &= \bigcup_{r\in\mathbb{Q}} [\{f > a - r\} \cap \{g > r\}], \qquad a \in \mathbb{R}, \\ fg &= \frac{1}{4}\left[(f+g)^2 - (f-g)^2\right]. \end{aligned}$$

Example 3. *It is* not *always true that $f \circ \Phi$ is measurable when $\Phi : \mathbb{R} \to \mathbb{R}$ is continuous and f is measurable. To see this recall the construction of the Cantor set $E \equiv \bigcap_{k=0}^{\infty} K_k$, where $K_k = \bigcup_{j=1}^{2^k} I_j^k$. Denote the open middle third of the closed* *interval* I_j^k *by* G_j^k. *Define the Cantor function* $F:[0,1]\to[0,1]$ *by*

$$\begin{aligned}
F(x) &= \frac{1}{2^1} \text{ for } x\in G_1^0=\left(\frac{1}{3},\frac{2}{3}\right);\\
F(x) &= \frac{1}{2^2} \text{ for } x\in G_1^1=\left(\frac{1}{9},\frac{2}{9}\right),\quad F(x)=\frac{3}{2^2} \text{ for } x\in G_2^1=\left(\frac{7}{9},\frac{8}{9}\right);\\
F(x) &= \frac{1}{2^3} \text{ for } x\in G_1^2,\quad F(x)=\frac{3}{2^3} \text{ for } x\in G_2^2,\\
&\qquad F(x)=\frac{5}{2^3} \text{ for } x\in G_3^2,\quad F(x)=\frac{7}{2^3} \text{ for } x\in G_4^2;\\
F(x) &= \frac{2j-1}{2^k} \text{ for } x\in G_j^{k-1},\qquad 1\le j\le 2^k,\qquad k\ge 1,
\end{aligned}$$

and then extend F *to the Cantor set* $E=[0,1]\setminus\left(\bigcup_{k,j}G_j^k\right)$ *by continuity. (*Exercise: *Prove there exists a unique continuous extension.) Now define*

$$G(x)=\frac{F(x)+x}{2},\qquad 0\le x\le 1.$$

Then $G:[0,1]\hookrightarrow[0,1]$ *is one-to-one (strictly increasing) and onto, hence the inverse function* $\Phi\equiv G^{-1}:[0,1]\hookrightarrow[0,1]$ *is continuous by Corollary 6. Now* $|G([0,1]\setminus E)|=\frac{1}{2}|[0,1]\setminus E|=\frac{1}{2}$ *by construction, and so* $|G(E)|=1-\frac{1}{2}=\frac{1}{2}$. *We have*

$$G(E)=\dot{\bigcup_{n\ge 1}}\{G(E)\cap(A\oplus r_n)\},$$

and if $B_n\equiv G(E)\cap(A\oplus r_n)\in\mathcal{L}$, *then*

$$\sum_{j=1}^{\infty}|B_n\oplus r_j|=\left|\dot{\bigcup_{j\ge 1}}(B_n\oplus r_j)\right|\le 1$$

implies that $|B_n|=0$. *Since* $|G(E)|>0$, *it follows that* $B_n\notin\mathcal{L}$ *for some* $n\ge 1$. *Denote such a set* B_n *by* B. *Then* $f=\chi_{\Phi(B)}$ *is measurable since* $\Phi(B)\subset E$ *is a null set. On the other hand,* $f\circ\Phi=\chi_B$ *is* not *meaurable, despite the continuity of* Φ.

Recall that a measurable simple function φ (i.e. the range of φ is finite) has the form

$$\varphi=\sum_{k=1}^{N}\alpha_k\chi_{E_k},\qquad \alpha_k\in\mathbb{R},E_k\in\mathcal{L}.$$

Next we collect two approximation properties of simple functions.

PROPOSITION 3. *Let* $f:\mathbb{R}\to[-\infty,\infty]$ *be measurable.*

(1) *If* f *is nonnegative there is an increasing sequence of nonnegative simple functions* $\{\varphi_k\}_{k=1}^{\infty}$ *that converges pointwise and monotonically to* f*:*

$$\varphi_k(x)\le\varphi_{k+1}(x) \text{ and } \lim_{k\to\infty}\varphi_k(x)=f(x),\qquad \text{for all } x\in\mathbb{R}.$$

(2) *There is a sequence of simple functions* $\{\varphi_k\}_{k=1}^{\infty}$ *satisfying*

$$|\varphi_k(x)|\le|\varphi_{k+1}(x)| \text{ and } \lim_{k\to\infty}\varphi_k(x)=f(x),\qquad \text{for all } x\in\mathbb{R}.$$

Comments: To prove (1) let $f_M = \min\{f, M\}$, and for $0 \le n < NM$ define

$$E_{n,N,M} = \left\{ x \in \mathbb{R} : \frac{n}{N} < f_M(x) \le \frac{n+1}{N} \right\}.$$

Then $\varphi_k(x) = \sum_{n=1}^{2^k k} \frac{n}{2^k} \chi_{E_{n,2^k,k}}(x)$ works. Property (2) follows from applying (1) to the positive and negative parts of f:

$$f^+(x) = \max\{f(x), 0\} \text{ and } f_-(x) = \max\{-f(x), 0\}.$$

2.2. Properties of integration and convergence theorems. If φ is a measurable simple function (i.e. its range is a finite set), then φ has a unique canonical representation

$$\varphi = \sum_{k=1}^{N} \alpha_k \chi_{E_k},$$

where the real constants α_k are distinct and nonzero, and the measurable sets E_k are pairwise disjoint. We define the Lebesgue integral of φ by

$$\int \varphi(x)\,dx = \sum_{k=1}^{N} \alpha_k |E_k|.$$

If E is a measurable subset of $\mathbb{R}$ and φ is a measurable simple function, then so is $\chi_E \varphi$, and we define

$$\int_E \varphi(x)\,dx = \int (\chi_E \varphi)(x)\,dx.$$

LEMMA 12. *Suppose that φ and ψ are measurable simple functions and that $E, F \in \mathcal{L}$.*

(1) *If $\varphi = \sum_{k=1}^{M} \beta_k \chi_{F_k}$ (not necessarily the canonical representation), then*

$$\int \varphi(x)\,dx = \sum_{k=1}^{M} \beta_k |F_k|.$$

(2) $\int (a\varphi + b\psi) = a \int \varphi + b \int \psi$ *for* $a, b \in \mathbb{C}$,
(3) $\int_{E \cup F} \varphi = \int_E \varphi + \int_F \varphi$ *if* $E \cap F = \phi$,
(4) $\int \varphi \le \int \psi$ *if* $\varphi \le \psi$,
(5) $\left| \int \varphi \right| \le \int |\varphi|$.

Properties (2) - (5) are usually referred to as *linearity*, *additivity*, *monotonicity* and the *triangle inequality* respectively. The proofs of (1) - (5) are routine.

Now we turn to defining the integral of a nonnegative measurable function $f : \mathbb{R} \to [0, \infty]$. For such f we define

$$\int f(x)\,dx = \sup \left\{ \int \varphi(x)\,dx : 0 \le \varphi \le f \text{ and } \varphi \text{ is simple} \right\}.$$

It is essential here that f be permitted to take on the value ∞, and that the supremum may be ∞ as well. We say that f is (Lebesgue) *integrable* if $\int f(x)\,dx < \infty$. For E measurable define

$$\int_E f(x)\,dx = \int (\chi_E f)(x)\,dx.$$

Here is an analogue of Lemma 12 whose proof is again routine.

Lemma 13. *Suppose that* $f, g : \mathbb{R} \to [0, \infty]$ *are nonnegative measurable functions and that* $E, F \in \mathcal{L}$.

(1) $\int (af + bg) = a \int f + b \int g$ *for* $a, b \in (0, \infty)$,
(2) $\int_{E \cup F} f = \int_E f + \int_F f$ *if* $E \cap F = \phi$,
(3) $\int f \leq \int g$ *if* $0 \leq f \leq g$,
(4) *If* $\int f < \infty$, *then* $f(x) < \infty$ *for a.e.* x,
(5) *If* $\int f = 0$, *then* $f(x) = 0$ *for a.e.* x.

Note that convergence of integrals does not always follow from pointwise convergence of the integrands. For example,

$$\lim_{n\to\infty} \int \chi_{[n,n+1]}(x)\,dx = 1 \neq 0 = \int \lim_{n\to\infty} \chi_{[n,n+1]}(x)\,dx,$$

and

$$\lim_{n\to\infty} \int n\chi_{\left(0,\frac{1}{n}\right)}(x)\,dx = 1 \neq 0 = \int \lim_{n\to\infty} n\chi_{\left[0,\frac{1}{n}\right]}(x)\,dx.$$

In each of these examples, the mass of the integrands "disappears" in the limit; at "infinity" in the first example and at the origin in the second example. Here are our first two classical convergence theorems giving conditions under which convergence *does* hold. The first generalizes the property in line 4 of (1.1):

$$\mu(E) = \sup_n \mu(E_n) = \lim_{n\to\infty} \mu(E_n) \text{ if } E_n \nearrow E.$$

Theorem 22. *(Monotone Convergence Theorem) Suppose that* $\{f_n\}_{n=1}^{\infty}$ *is an* increasing *sequence of nonnegative measurable functions, i.e.* $f_n(x) \leq f_{n+1}(x)$, *and let*

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

Then f *is nonegative and measurable and*

$$\int f(x)\,dx = \sup_n \int f_n(x)\,dx = \lim_{n\to\infty} \int f_n(x)\,dx.$$

Proof: Since $\int f_n \leq \int f_{n+1}$ we have $\lim_{n\to\infty} \int f_n = L \in [0, \infty]$. Now f is measurable and $f_n \leq f$ implies $\int f_n \leq \int f$ so that

$$L \leq \sup_n \int f_n \leq \int f.$$

To prove the opposite inequality, momentarily fix a simple function φ such that $0 \leq \varphi \leq f$. Choose $c < 1$ and define

$$E_n = \{x \in \mathbb{R} : f_n(x) \geq c\varphi(x)\}, \qquad n \geq 1.$$

Then E_n is an increasing sequence of measurable sets with $\bigcup_{n=1}^{\infty} E_n = \mathbb{R}$. We have

$$\int f_n \geq \int_{E_n} f_n \geq c \int_{E_n} \varphi, \qquad n \geq 1.$$

Now let $\varphi = \sum_{k=1}^{N} \alpha_k \chi_{F_k}$ be the canonical representation of φ. Then

$$\int_{E_n} \varphi = \sum_{k=1}^{N} \alpha_k |E_n \cap F_k|,$$

and since $\lim_{n\to\infty} |E_n \cap F_k| = |F_k|$ by the fourth line in (1.1), we obtain that

$$\int_{E_n} \varphi = \sum_{k=1}^{N} \alpha_k |E_n \cap F_k| \to \sum_{k=1}^{N} \alpha_k |F_k| = \int \varphi$$

as $n \to \infty$. Altogether then we have

$$L = \lim_{n\to\infty} \int f_n \geq c \int \varphi$$

for all $c < 1$, which implies $L \geq \int \varphi$ for all simple φ with $0 \leq \varphi \leq f$, which implies $L \geq \int f$ as required.

Note that as a corollary we have $\int f = \lim_{k\to\infty} \int \varphi_k$ where the simple functions φ_k are as in (1) of Proposition 3. We also have this.

COROLLARY 9. *Suppose that $a_k(x) \geq 0$ is measurable for $k \geq 1$. Then*

$$\int \sum_{k=1}^{\infty} a_k(x)\,dx = \sum_{k=1}^{\infty} \int a_k(x)\,dx.$$

To prove the corollary apply the Monotone Convergence Theorem to the sequence of partial sums $f_n(x) = \sum_{k=1}^{n} a_k(x)$.

LEMMA 14. *(Fatou's Lemma) If $\{f_n\}_{n=1}^{\infty}$ is a sequence of* nonnegative *measurable functions, then*

$$\int \liminf_{n\to\infty} f_n(x)\,dx \leq \liminf_{n\to\infty} \int f_n(x)\,dx.$$

Proof: Let $g_n(x) = \inf_{k\geq n} f_k(x)$ so that $g_n \leq f_n$ and $\int g_n \leq \int f_n$. Then $\{g_n\}_{n=1}^{\infty}$ is an increasing sequence of nonnegative measurable functions that converges pointwise to $\liminf_{n\to\infty} f_n(x)$. So the Monotone Convergence Theorem yields

$$\int \liminf_{n\to\infty} f_n(x)\,dx = \lim_{n\to\infty} \int g_n(x)\,dx \leq \liminf_{n\to\infty} \int f_n(x)\,dx.$$

Finally, we can give an unambiguous meaning to the integral $\int f(x)\,dx$ in the case when f is *integrable*, by which we mean that f is measurable and $\int |f(x)|\,dx < \infty$. To do this we note that the positive and negative parts of f,

$$f^+(x) = \max\{f(x), 0\} \text{ and } f_-(x) = \max\{-f(x), 0\},$$

are both nonnegative measurable functions with finite integral. We define

$$\int f(x)\,dx = \int f^+(x)\,dx - \int f_-(x)\,dx.$$

With this definition we have the usual elementary properties of linearity, additivity, monotonicity and the triangle inequality.

LEMMA 15. *Suppose that f, g are integrable and that $E, F \in \mathcal{L}$.*

(1) $\int (af + bg) = a\int f + b\int g$ *for* $a, b \in \mathbb{R}$,
(2) $\int_{E\cup F} f = \int_E f + \int_F f$ *if* $E \cap F = \phi$,
(3) $\int f \leq \int g$ *if* $f \leq g$,
(4) $\left|\int f\right| \leq \int |f|$.

Our final convergence theorem is one of the most useful in analysis.

THEOREM 23. *(Dominated Convergence Theorem) Let g be a nonnegative integrable function. Suppose that $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions satisfying*

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad a.e.\ x,$$

and

$$|f_n(x)| \leq g(x), \qquad a.e.\ x.$$

Then

$$\lim_{n\to\infty} \int |f(x) - f_n(x)|\, dx = 0,$$

and hence

$$\int f(x)\, dx = \lim_{n\to\infty} \int f_n(x)\, dx.$$

Proof: Since $|f| \leq g$ and f is measurable, f is integrable. Since $|f - f_n| \leq 2g$, Fatou's Lemma can be applied to the sequence of functions $2g - |f - f_n|$ to obtain

$$\begin{aligned} \int 2g &\leq \liminf_{n\to\infty} \int (2g - |f - f_n|) \\ &= \int 2g + \liminf_{n\to\infty} \left(-\int |f - f_n| \right) \\ &= \int 2g - \limsup_{n\to\infty} \int |f - f_n|. \end{aligned}$$

Since $\int 2g < \infty$, we can subtract it from both sides to obtain

$$\limsup_{n\to\infty} \int |f - f_n| \leq 0,$$

which implies $\lim_{n\to\infty} \int |f - f_n| = 0$. Then $\int f = \lim_{n\to\infty} \int f_n$ follows from the triangle inequality $\left| \int (f - f_n) \right| \leq \int |f - f_n|$.

Note that as a corollary we have $\int f = \lim_{k\to\infty} \int \varphi_k$ where the simple functions φ_k are as in (2) of Proposition 3.

Finally, if $f(x) = u(x) + iv(x)$ is complex-valued where $u(x)$ and $v(x)$ are real-valued measurable functions such that

$$\int |f(x)|\, dx = \int \sqrt{u(x)^2 + v(x)^2}\, dx < \infty,$$

then we define

$$\int f(x)\, dx = \int u(x)\, dx + i \int v(x)\, dx.$$

The usual properties of linearity, additivity, monotonicity and the triangle inequality all hold for this definition as well.

2.3. Three famous measure problems. The following three problems are listed in order of increasing difficulty.

PROBLEM 1. *Suppose that* $E_1, ..., E_n$ *are* n *Lebesgue measurable subsets of* $[0,1]$ *such that each point* x *in* $[0,1]$ *lies in some* k *of these subsets. Prove that there is at least one set* E_j *with* $|E_j| \geq \frac{k}{n}$.

PROBLEM 2. *Suppose that* E *is a Lebesgue measurable set of positive measure. Prove that*

$$E - E = \{x - y : x, y \in E\}$$

contains a nontrivial open interval.

PROBLEM 3. *Construct a Lebesgue measurable subset of the real line such that*

$$0 < \frac{|E \cap I|}{|I|} < 1$$

for all nontrivial open intervals I.

To solve Problem 1, note that the hypothesis implies $k \leq \sum_{j=1}^{n} \chi_{E_j}(x)$ for $x \in [0,1]$. Now integrate to obtain

$$k = \int_0^1 k dx \leq \int_0^1 \left(\sum_{j=1}^{n} \chi_{E_j}(x) \right) dx = \sum_{j=1}^{n} \int_0^1 \chi_{E_j}(x) \, dx = \sum_{j=1}^{n} |E_j| \, ,$$

which implies that $|E_j| \geq \frac{k}{n}$ for some j. The solution is much less elegant without recourse to integration.

To solve Problem 2, choose K compact contained in E such that $|K| > 0$. Then choose G open containing K such that $|G \setminus K| < |K|$. Let $\delta = dist(K, G^c) > 0$. It follows that $(-\delta, \delta) \subset K - K \subset E - E$. Indeed, if $x \in (-\delta, \delta)$ then $K - x \subset G$ and $K \cap (K - x) \neq \phi$ since otherwise we have a contradiction:

$$2|K| = |K| + |K - x| \leq |G| \leq |G \setminus K| + |K| < 2|K| .$$

Thus there are k_1 and k_2 in K such that $k_1 = k_2 - x$ and so

$$x = k_2 - k_1 \in K - K.$$

Problem 3 is most easily solved using generalized Cantor sets E_α. Let $0 < \alpha \leq 1$ and set $I_1^0 = [0,1]$. Remove the open interval of length $\frac{1}{3}\alpha$ centered in I_1^0 and denote the two remaining closed intervals by I_1^1 and I_2^1. Then remove the open interval of length $\frac{1}{3^2}\alpha$ centered in I_1^1 and denote the two remaining closed intervals by I_1^2 and I_2^2. Do the same for I_2^1 and denote the two remaining closed intervals by I_3^2 and I_4^2.

Continuing in this way, we obtain at the k^{th} generation, a collection $\left\{I_j^k\right\}_{j=1}^{2^k}$ of 2^k pairwise disjoint closed intervals of equal length. Let

$$E_\alpha = \bigcap_{k=1}^{\infty} \left(\bigcup_{j=1}^{2^k} I_j^k \right).$$

Then by summing the lengths of the removed open intervals, we obtain

$$|[0,1] \setminus E_\alpha| = \frac{1}{3}\alpha + \frac{2}{3^2}\alpha + \frac{2^2}{3^3}\alpha + ... = \alpha,$$

and it follows that E_α is compact and has Lebesgue measure $1-\alpha$. It is not hard to show that E_α is also nowhere dense. The case $\alpha = 1$ is particularly striking: E_1 is a compact, perfect and uncountable subset of $[0,1]$ having Lebesgue measure 0. This is the classical Cantor set introduced at the end of Chapter 1.

In order to construct the set E in Problem 3, it suffices by taking unions of translates by integers, to construct a subset E of $[0,1]$ satisfying

$$0 < \frac{|E \cap I|}{|I|} < 1, \qquad \text{for all intervals } I \subset [0,1] \text{ of positive length.} \tag{2.1}$$

Fix $0 < \alpha_1 < 1$ and start by taking $E^1 = E_{\alpha_1}$. It is not hard to see that $\frac{|E^1 \cap I|}{|I|} < 1$ for all I, but the left hand inequality in (2.1) fails for $E = E^1$ whenever I is a subset of one of the component intervals in the open complement $[0,1] \setminus E^1$. To remedy this fix $0 < \alpha_2 < 1$ and for each component interval J of $[0,1] \setminus E^1$, translate and dilate E_{α_2} to fit snugly in the closure $\overline{J}$ of the component, and let E^2 be the union of E^1 and all these translates and dilates of E_{α_2}. Then again, $\frac{|E^2 \cap I|}{|I|} < 1$ for all I but the left hand inequality in (2.1) fails for $E = E^2$ whenever I is a subset of one of the component intervals in the open complement $[0,1] \setminus E^2$. Continue this process indefinitely with a sequence of numbers $\{\alpha_n\}_{n=1}^\infty \subset (0,1)$. We claim that $E = \bigcup_{n=1}^\infty E^n$ satisfies (2.1) if and only if

$$\sum_{n=1}^\infty (1-\alpha_n) < \infty. \tag{2.2}$$

To see this, first note that no matter what sequence of numbers α_n *less than one* is used, we obtain that $0 < \frac{|E \cap I|}{|I|}$ for all intervals I of positive length. Indeed, each set E^n is easily seen to be compact and nowhere dense, and each component interval in the complement $[0,1] \setminus E^n$ has length at most

$$\frac{\alpha_1}{3}\frac{\alpha_2}{3}...\frac{\alpha_n}{3} \leq 3^{-n}.$$

Thus given an interval I of positive length, there is n large enough such that I will contain one of the component intervals J of $[0,1] \setminus E^n$, and hence will contain the translated and dilated copy $\mathcal{C}\left(E_{\alpha_{n+1}}\right)$ of $E_{\alpha_{n+1}}$ that is fitted into J by construction. Since the dilation factor is the length $|J|$ of J, we have

$$|E \cap I| \geq \left|\mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| = |J| \left|E_{\alpha_{n+1}}\right| = |J| (1-\alpha_{n+1}) > 0,$$

since $\alpha_{n+1} < 1$.

It remains to show that $|E \cap I| < |I|$ for all intervals I of positive length in $[0,1]$, and it is here that we must use (2.2). Indeed, fix I and let J be a component interval of $[0,1] \setminus E^n$ (with n large) that is contained in I. Let $\mathcal{C}\left(E_{\alpha_{n+1}}\right)$ be the translated and dilated copy of $E_{\alpha_{n+1}}$ that is fitted into J by construction. We compute that

$$\begin{aligned} |E \cap J| &= \left|\mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| + (1-\alpha_{n+2}) \left|J \setminus \mathcal{C}\left(E_{\alpha_{n+1}}\right)\right| + ... \\ &= (1-\alpha_{n+1}) |J| + (1-\alpha_{n+2}) \alpha_{n+1} |J| \\ &\quad + (1-\alpha_{n+3}) \alpha_{n+2}\alpha_{n+1} |J| + ... \\ &= \sum_{k=1}^\infty \beta_k^n |J|, \end{aligned}$$

where

$$\beta_k^n = (1-\alpha_{n+k})\,\alpha_{n+k-1}...\alpha_{n+1}, \qquad k \geq 1.$$

Then we have

$$|E \cap J| = \left(\sum_{k=1}^{\infty} \beta_k^n\right) |J| < |J|,$$

and hence also $\frac{|E\cap I|}{|I|} < 1$, if we choose $\{\alpha_n\}_{n=1}^{\infty}$ so that $\sum_{k=1}^{\infty} \beta_k^n < 1$ for all n.

Now by induction we have

$$\sum_{k=1}^{\infty} \beta_k^n = \lim_{N\to\infty} \sum_{k=1}^{N} (1-\alpha_{n+k})\,\alpha_{n+k-1}...\alpha_{n+1} = \lim_{N\to\infty}\left(1-\prod_{k=1}^{N}\alpha_{n+k}\right) = 1-\prod_{k=1}^{\infty}\alpha_{n+k},$$

and by the first line in (2.3) below, this is strictly less than 1 *if and only if* $\sum_{k=1}^{\infty}(1-\alpha_{n+k}) < \infty$ for all n. Thus the set E constructed above satisfies (2.1) if and only if (2.2) holds.

2.3.1. *Infinite products.* If $0 \leq u_n < 1$ and $0 \leq v_n < \infty$ then

$$\begin{aligned} \prod_{n=1}^{\infty}(1-u_n) &> 0 \text{ if and only if } \sum_{n=1}^{\infty} u_n < \infty, \\ \prod_{n=1}^{\infty}(1+v_n) &< \infty \text{ if and only if } \sum_{n=1}^{\infty} v_n < \infty. \end{aligned} \tag{2.3}$$

To see (2.3) we may assume $0 \leq u_n, v_n \leq \frac{1}{2}$, so that $e^{-u_n} \geq 1-u_n \geq e^{-2u_n}$ and $e^{\frac{1}{2}v_n} \leq 1+v_n \leq e^{v_n}$. For example, when $0 \leq x \leq \frac{1}{2}$, the alternating series estimate yields

$$e^{-2x} \leq 1-2x+\frac{(2x)^2}{2!} \leq 1-x,$$

while the geometric series estimate yields

$$e^{\frac{1}{2}x} \leq 1+\left(\frac{1}{2}x\right)\left\{1+x+x^2+...\right\} \leq 1+x.$$

Thus we have

$$\begin{aligned} \exp\left(-\sum_{n=1}^{\infty}u_n\right) &\geq \prod_{n=1}^{\infty}(1-u_n) \geq \exp\left(-2\sum_{n=1}^{\infty}u_n\right), \\ \exp\left(\frac{1}{2}\sum_{n=1}^{\infty}v_n\right) &\leq \prod_{n=1}^{\infty}(1+v_n) \leq \exp\left(\sum_{n=1}^{\infty}v_n\right). \end{aligned} \tag{2.4}$$

CHAPTER 5

Paradoxical decompositions and finitely additive measures

DEFINITION 12. *Let G be a group acting on a set X. A subset E of X is* finitely G-paradoxical *if there are subsets A_i, B_j of X and group elements g_i, h_j such that*

$$
\begin{aligned}
E &\supset \left(\dot\cup_{i=1}^{m} A_i\right) \dot\cup \left(\dot\cup_{j=1}^{n} B_j\right), \\
E &= \cup_{i=1}^{m} g_i A_i = \cup_{j=1}^{n} h_j B_j .
\end{aligned} \tag{0.5}
$$

The notation $\dot\cup$ asserts that the indicated union is pairwise disjoint. Note that one can easily arrange to have each collection of sets $\{g_i A_i\}_{i=1}^{m}$ and $\{h_j B_j\}_{j=1}^{n}$ in the second line of (0.5) pairwise disjoint simply by paring the sets A_i and B_j. One can also achieve equality in the first line of (0.5), but this is harder, and is not proved until Corollary 12 below. We say that E is *countably G-paradoxical* if m, n in (0.5) are permitted to be ∞, the first infinite ordinal. By G-paradoxical we mean finitely G-paradoxical. Finally, we say that G is paradoxical if G acts on itself by left multiplication and G is G-paradoxical. The next result uses the axiom of choice.

THEOREM 24. *Let G be the circle group $\mathbb{T}$ and let it act on itself $X = \mathbb{T}$ by group multiplication:*

$$e^{it} \in G \text{ sends the point } e^{ix} \in X \text{ to the point } e^{i(t+x)} \in X.$$

Then X is countably G-paradoxical.

Proof: Let M be a choice set for the equivalence classes of the relation on $\mathbb{T}$ given by declaring two points equivalent if one is obtained from the other by rotation through a rational multiple of 2π radians. Let $\{\rho_i\}_{i=1}^{\infty}$ enumerate the rotations through a rational multiple of 2π radians, and set $M_i = \rho_i M$. Then the countable paradoxical decomposition is provided by

$$
\begin{aligned}
X &= \left(\dot\cup_{i \text{ odd}} M_i\right) \dot\cup \left(\dot\cup_{i \text{ even}} M_i\right), \\
X &= \dot\cup_{i \text{ odd}} g_i M_i = \dot\cup_{i \text{ even}} h_i M_i,
\end{aligned}
$$

where $g_i = \rho_{\frac{i+1}{2}} \rho_i^{-1}$ for i odd, and $h_i = \rho_{\frac{i}{2}} \rho_i^{-1}$ for i even.

COROLLARY 10. *There is a non-Lebesgue measurable subset of $\mathbb{T}$.*

Proof: If A_i, B_j, g_i, h_j witness a countable paradoxical decomposition (0.5) of $\mathbb{T} = E$ with $m, n \leq \infty$, and if we assume every subset of $\mathbb{T}$ is Lebesgue measurable, then

$$
\begin{aligned}
2\pi &= |G| \geq \sum_{i=1}^{m} |A_i| + \sum_{j=1}^{n} |B_j| = \sum_{i=1}^{m} |g_i A_i| + \sum_{j=1}^{n} |h_j B_j| \\
&\geq \left|\cup_{i=1}^{m} g_i A_i\right| + \left|\cup_{j=1}^{n} h_j B_j\right| = 4\pi,
\end{aligned}
$$

a contradiction.

Denote by G_n the group of *isometries* of Euclidean space $\mathbb{R}^n$.

REMARK 10. *There exists a G_2-paradoxical subset E of the plane $\mathbb{R}^2 = \mathbb{C}$ that does* not *require the axiom of choice for its construction, namely the Sierpiński-Mazurkiewicz Paradox: let $e^{i\theta}$ be a transcendental complex number and define*

$$\begin{aligned} E &= \left\{ x = \sum_{n=0}^{\infty} x_n e^{in\theta} \in \mathbb{C} : x_n \in \mathbb{Z}_+ \text{ and } x_n = 0 \text{ for all but finitely many } n \right\}, \\ E_1 &= \{x \in E : x_0 = 0\}, \\ E_2 &= \{x \in E : x_0 > 0\}. \end{aligned}$$

Then $E = E_1 \dot{\cup} E_2 = e^{-i\theta} E_1 = E_2 - 1$.

1. Finitely additive invariant measures

Let G be a group acting on a set X. If there exists a finitely (countably) additive G-invariant probability measure μ on the power set $\mathcal{P}(X)$, then there are no finitely (countably) G-paradoxical subsets E of X having positive μ-measure. In particular G itself is not finitely (countably) G-paradoxical. This is proved as in the proof of Corollary 10 above. Thus paradoxical constructions can be viewed as nonexistence theorems for invariant measures, and by the contrapositive, the construction of invariant measures precludes paradoxical decompositions. In fact a theorem of Tarski shows that if $E \subset X$ on which a group acts, then there is a finitely additive G-invariant positive measure μ on $\mathcal{P}(X)$ with $\mu(E) = 1$ *if and only if* E is not G-paradoxical.

We now state two theorems in this regard. The first states that paradoxical decompositions *never* occur for abelian groups (such as the group of translations on Euclidean space $\mathbb{R}^n$), and the second shows that paradoxical decompositions *do* exist for the rotation groups on Euclidean space $\mathbb{R}^n$ when $n \geq 3$ (resulting in the Banach-Tarski paradox).

THEOREM 25. *Suppose G is an abelian group and let $\mathcal{M}$ be the power set of G. There is $\mu : \mathcal{M} \to [0,1]$ satisfying*

(1) $\mu(E_1 \dot{\cup} E_2) = \mu(E_1) + \mu(E_2), \qquad E_i \in \mathcal{M}$,
(2) $\mu(E + a) = \mu(E), \qquad E \in \mathcal{M}, a \in G$,
(3) $\mu(G) = 1$.

DEFINITION 13. *Let G act on a set X. Subsets A and B of X are said to be G-equidecomposable, written $A \sim_G B$ or simply $A \sim B$ when G is understood, if $A = \dot{\cup}_{i=1}^n A_i$ and $B = \dot{\cup}_{i=1}^n B_i$ where $A_i = g_i B_i$ for some $g_i \in G$, $1 \leq i \leq n$.*

We will see later that E is G-paradoxical if and only if $E = A \dot{\cup} B$ where $A \sim_G E \sim_G B$.

REMARK 11. *If X is Euclidean space $\mathbb{R}^n$, then G_3-equidecomposability preserves the following properties: boundedness, Lebesgue measure zero, first category (a countable union of nowhere dense sets), and second category (not first category).*

THEOREM 26. *(Banach-Tarski paradox) The sphere $\mathbb{S}^2$ is SO_3-paradoxical and the ball $\mathbb{B}_3$ is G_3-paradoxical. Moreover, if A and B are any two bounded subsets of $\mathbb{R}^3$, each having nonempty interior, then A and B are G_3-equidecomposable.*

We prove only the second theorem on the Banach-Tarski paradox.

2. Paradoxical decompositions and the Banach-Tarski paradox

We obtain the strong form of the Banach-Tarski paradox in four steps.

- First, we prove that the free nonabelian group F_2 of rank 2 is paradoxical.
- Second, we show that the special orthogonal group SO_3 in three dimensions contains a copy of F_2.
- Third, we lift the paradoxical decomposition from SO_3 to the sphere $\mathbb{S}^2$ on which it acts "almost" without nontrivial fixed points.
- Fourth, we extend the paradox to bounded sets with nonempty interior with the help of the proof of the Schröder-Bernstein theorem.

First step: We prove that F_2 is paradoxical. Let F_2 consist of all finite "words" in $\sigma, \sigma^{-1}, \tau, \tau^{-1}$ with concatenation as the group operation, and the empty word as identity 1. For $\rho \in \left\{\sigma, \sigma^{-1}, \tau, \tau^{-1}\right\}$, let $W(\rho)$ consist of all reduced words that begin with ρ (a word is *reduced* if no pair of adjacent symbols is $\sigma\sigma^{-1}$, $\sigma^{-1}\sigma$, $\tau\tau^{-1}$, or $\tau^{-1}\tau$). The following decompositions witness the paradoxical nature of F_2:

$$\begin{aligned}
F_2 &= \{1\} \dot{\cup} W(\sigma) \dot{\cup} W\left(\sigma^{-1}\right) \dot{\cup} W(\tau) \dot{\cup} W\left(\tau^{-1}\right), \\
F_2 &= W(\sigma) \dot{\cup} \sigma W\left(\sigma^{-1}\right), \\
F_2 &= W(\tau) \dot{\cup} \tau W\left(\tau^{-1}\right).
\end{aligned}$$

Note that we do not use the identity in these reconstructions of F_2. We can however witness the paradox with four disjoint pieces whose union is F_2 using an *absorption process* as follows. First we include 1 with the set $W(\sigma)$ and call the new set A_1. But then $F_2 = A_1 \dot{\cup} \sigma W\left(\sigma^{-1}\right)$ fails since 1 is also in $\sigma W\left(\sigma^{-1}\right)$. So 1 must be removed from $\sigma W\left(\sigma^{-1}\right)$, and we achieve this by moving σ^{-1} from $W\left(\sigma^{-1}\right)$ to A_1 and denoting the new set $W\left(\sigma^{-1}\right) \setminus \left\{\sigma^{-1}\right\}$ by A_2. But then σ^{-1} is in both A_1 and A_2. So we move σ^{-2} from A_2 to A_1. This process must be continued indefinitely, so let $S = \{\sigma^{-n}\}_{n=1}^{\infty}$ and define

$$\begin{aligned}
A_1 &= W(\sigma) \dot{\cup} \{1\} \dot{\cup} S, \\
A_2 &= W\left(\sigma^{-1}\right) \setminus S, \\
A_3 &= W(\tau), \\
A_4 &= W\left(\tau^{-1}\right).
\end{aligned}$$

Then $F_2 = \dot{\cup}_{i=1}^{4} A_i$ and $F_2 = A_1 \dot{\cup} \sigma A_2$ and $F_2 = A_3 \dot{\cup} \tau A_4$ since

$$\begin{aligned}
\sigma A_2 &= \sigma W\left(\sigma^{-1}\right) \setminus \sigma S = \left\{\{1\} \dot{\cup} W\left(\sigma^{-1}\right) \dot{\cup} W(\tau) \dot{\cup} W\left(\tau^{-1}\right)\right\} \setminus \{\{1\} \dot{\cup} S\} \\
&= \left\{W\left(\sigma^{-1}\right) \setminus S\right\} \dot{\cup} W(\tau) \dot{\cup} W\left(\tau^{-1}\right),
\end{aligned}$$

has complement A_1.

Second step: To embed a copy of F_2 in SO_3 we define the 3×3 matrices:

$$\begin{aligned}
\phi^{\pm} &= \begin{bmatrix} \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} & 0 \\ \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & \mp 2\sqrt{2} & 0 \\ \pm 2\sqrt{2} & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \\
\rho^{\pm} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & \mp\frac{2\sqrt{2}}{3} \\ 0 & \pm\frac{2\sqrt{2}}{3} & \frac{1}{3} \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 3 & 0 & 0 \\ 0 & 1 & \mp 2\sqrt{2} \\ 0 & \pm 2\sqrt{2} & 1 \end{bmatrix}.
\end{aligned}$$

It suffices to show that no nonempty reduced word w in $\phi^{\pm}, \rho^{\pm}$ equals the identity in SO_3. Since conjugation by $\phi^{\pm}$ doesn't affect the vanishing of a word, we may assume that w is a nonempty reduced word ending in $\phi^{\pm}$.

Claim 4. *Every nonempty reduced word w in $\phi^{\pm}, \rho^{\pm}$ that ends in $\phi^{\pm}$ satisfies $w(1,0,0) = 3^{-k}\left(a, b\sqrt{2}, c\right)$ for some $a, b, c \in \mathbb{Z}$ with $3 \nmid b$, and where k is the length of w. In particular $b \neq 0$ and w is* not *the identity.*

We prove the claim by induction on the length k of w. The case $k = 1$ is evident upon examining the first columns of the two matrices $\phi^{\pm}$. If w of length $k \geq 2$ equals $\phi^{\pm} w'$ or $\rho^{\pm} w'$, where

$$w'(1,0,0) = 3^{1-k}\left(a', b'\sqrt{2}, c'\right), \qquad a', b', c' \in \mathbb{Z},\ 3 \nmid b',$$

then

$$\begin{aligned} \phi^{\pm} w'(1,0,0) &= 3^{-k}\left(a' \mp 4b', (b' \pm 2a')\sqrt{2}, 3c'\right), \\ \rho^{\pm} w'(1,0,0) &= 3^{-k}\left(3a', (b' \mp 2c')\sqrt{2}, c' \pm 4b'\right). \end{aligned} \tag{2.1}$$

We now see that $w(1,0,0)$ has the form $3^{-k}\left(a, b\sqrt{2}, c\right)$ for some $a, b, c \in \mathbb{Z}$, and it remains only to prove $3 \nmid b$ given that $3 \nmid b'$. There are four cases: $w = \phi^{\pm}\rho^{\pm}v$, $\rho^{\pm}\phi^{\pm}v$, $\phi^{\pm}\phi^{\pm}v$ and $\rho^{\pm}\rho^{\pm}v$ where v is possibly empty. We may suppose that $v(1,0,0) = 3^{2-k}\left(a'', b''\sqrt{2}, c''\right)$ where $a'', b'', c'' \in \mathbb{Z}$ (we do not assume $3 \nmid b''$ in order to include the case v is empty). In the first case, we have $a' = 3a''$ by the second line in (2.1) applied to v instead of w'. Now $3 \nmid b'$ and so we obtain $3 \nmid b' \pm 2a' = b$ as required. The second case is similar. For the third case we have

$$\begin{aligned} a' &= a'' - 4b'', \\ b' &= 2a'' + b'', \end{aligned}$$

by the first line in (2.1) applied to v instead of w'. Then

$$b = b' \pm 2a' = b' \pm 2\left(a'' \mp 4b''\right) = b' + b'' \pm 2a'' - 9b'' = 2b' - 9b'',$$

and again $3 \nmid b$ follows from $3 \nmid b'$. The fourth case is similar and this completes the proof of the claim.

Third step: To lift a paradoxical decomposition from a group to a set on which it acts is easy using the axiom of choice provided the action is with *trivial* fixed points. We say that a group G acts on a set X with *trivial* fixed points if $gx \neq x$ for all $x \in X$ and all $g \in G \setminus \{e\}$ where e denotes the identity element of G.

Proposition 4. *If G is a paradoxical group and acts on a set X with trivial fixed points, then X is G-paradoxical.*

Proof: Let A_i, B_j, g_i, h_j witness the paradoxical nature of G as in (0.5). Let M be a choice set for the G-orbits in X. Then $\{gM\}_{g\in G}$ is a partition of X because there are no nontrivial fixed points. Then $A_i^* = \dot\cup_{g\in A_i} gM$ and $B_j^* = \dot\cup_{h\in B_j} hM$ easily yield a paradoxical decomposition of X:

$$\begin{aligned} X &\supset \left(\dot\cup_{i=1}^m A_i^*\right) \dot\cup \left(\dot\cup_{j=1}^n B_j^*\right), \\ X &= \dot\cup_{i=1}^m g_i A_i^* = \dot\cup_{j=1}^n h_j B_j^*. \end{aligned}$$

Corollary 11. *(Hausdorff's paradox) There is a countable set $D \subset \mathbb{S}^2$ such that $\mathbb{S}^2 \diagdown D$ is SO_3-paradoxical.*

Proof: Let F be a free nonabelian group of rank 2 in SO_3. Then F is countable and since each $\alpha \in F \setminus \{1\}$ fixes exactly 2 points,

$$D = \left\{x \in \mathbb{S}^2 : \alpha x = x \text{ for some } \alpha \in F \setminus \{1\}\right\}$$

is countable. Then F acts on $\mathbb{S}^2 \diagdown D$ with trivial fixed points. Indeed, the set D of trivial fixed points of F is invariant for F since if $\xi \in D$ and $\alpha\xi = \xi$, then $\left(\beta\alpha\beta^{-1}\right)\beta\xi = \beta\xi$ for all $\alpha, \beta \in F$; and thus $\mathbb{S}^2 \diagdown D$ is invariant for F as well. So Proposition 4 implies that $\mathbb{S}^2 \diagdown D$ is F-paradoxical, hence also SO_3-paradoxical.

Hausdorff's paradox is already sufficient to disprove the existence of *finitely* additive rotation invariant positive measures of total mass 1 on the power set of $\mathbb{S}^2$, and hence also disproves the existence of *finitely* additive isometry invariant positive measures on the power set of $\mathbb{R}^3$ that normalize the unit cube (this was Hausdorff's motivation). *Exercise*: prove this! We can eliminate the countable set D in Hausdorff's paradox by an absorption process once we have the following lemma.

Lemma 16. *Let G act on a set X and let $E, E' \in \mathcal{P}(X)$. If $E \sim_G E'$, then E is G-paradoxical* if and only if *E' is G-paradoxical.*

First we note that the relation $\sim_G$ is transitive. Suppose that $E \sim_G A$ and $E \sim_G B$. Then $E = \dot\bigcup_{i=1}^n A_i = \dot\bigcup_{j=1}^m B_j$ where $A = \dot\bigcup_{i=1}^n g_i A_i$ and $B = \dot\bigcup_{j=1}^m h_j B_j$ for some group elements g_i, h_j. Then $A = \dot\bigcup_{i,j=1}^{n,m} g_i\left(A_i \cap B_j\right)$ and $B = \dot\bigcup_{i,j=1}^{n,m} h_j\left(A_i \cap B_j\right)$ shows that $A \sim_G B$. From this we easily obtain the lemma. Indeed, E is G-paradoxical if and only if there are disjoint subsets B_1, B_2 of E such that both $B_1 \sim_G E$ and $B_2 \sim_G E$. From $E \sim_G E'$, we have $E = \dot\bigcup_{i=1}^n A_i$ and $E' = \dot\bigcup_{i=1}^n g_i A_i$. Thus if we define $B_1' = \dot\bigcup_{i=1}^n g_i\left(A_i \cap B_1\right)$ and $B_2' = \dot\bigcup_{i=1}^n g_i\left(A_i \cap B_2\right)$, we have that B_1', B_2' are disjoint subsets of E' such that $B_1' \sim_G \dot\bigcup_{i=1}^n\left(A_i \cap B_1\right) = B_1 \sim_G E \sim_G E'$ and similarly $B_2' \sim_G E'$. This shows that E' is G-paradoxical.

Theorem 27. *(Banach-Tarski paradox) $\mathbb{S}^2$ is SO_3-paradoxical and $\mathbb{B}_3$ is G_3-paradoxical.*

Proof: Let $D = \{d_i\}_{i=1}^\infty$ be as in Hausdorff's paradox. Pick a line ℓ through the origin that misses D and fix a plane P containing ℓ. Let

$$A = \left\{\frac{1}{n}\left(\theta_i - \theta_j\right) : n, i, j \in \mathbb{N}\right\}, \qquad \theta_i = \measuredangle\left(d_i, \ell\right),$$

where $\measuredangle\left(d_i, \ell\right)$ denotes the angle mod π through which the plane P must be rotated (in a fixed sense) about ℓ so as to contain d_i. Pick $\theta \notin A$ (mod π). Then if ρ is rotation about ℓ through angle θ, we have

$$\rho^m D \cap \rho^n D = \phi, \qquad m \neq n \text{ in } \mathbb{Z}.$$

Indeed, if ℓ is the z-axis and $\rho^m D \cap \rho^n D \neq \phi$ for some $m \neq n$, then using polar coordinates in the xy-plane we have $e^{im\theta} r_j e^{i\theta_j} = e^{in\theta} r_k e^{i\theta_k}$, which implies $\theta = \frac{\theta_k - \theta_j}{m-n} \in A$. So with $\widetilde{D} = \dot\bigcup_{n=0}^\infty \rho^n D = D \dot\cup \rho\widetilde{D}$ we have

$$\mathbb{S}^2 = \left(\mathbb{S}^2 \diagdown \widetilde{D}\right) \dot\cup \widetilde{D} \sim_{SO_3} \left(\mathbb{S}^2 \diagdown \widetilde{D}\right) \dot\cup \rho\widetilde{D} = \mathbb{S}^2 \diagdown D,$$

and the lemma shows that $\mathbb{S}^2$ is SO_3-paradoxical.

Finally, the equality

$$\mathbb{B}_3 \diagdown \{0\} = \cup_{\omega \in S^2} \{\lambda\omega : 0 < \lambda < 1\}$$

shows that $\mathbb{B}_3 \diagdown \{0\}$ is SO_3-paradoxical, and an absorption argument as above then shows that $\mathbb{B}_3$ is G_3-paradoxical. Indeed, use a rotation ρ about a line ℓ passing through $\left(0, 0, \frac{1}{3}\right)$ but *not* passing through the origin, so that $\rho^m 0 \neq \rho^n 0$ for $m \neq n$, and set $\widetilde{D} = \dot{\cup}_{n=0}^{\infty} \rho^n 0 = \{0\} \dot{\cup} \rho\widetilde{D}$. Then since $\rho \in G_3$,

$$\mathbb{B}_3 = \left(\mathbb{B}_3 \diagdown \widetilde{D}\right) \dot{\cup} \widetilde{D} \sim_{G_3} \left(\mathbb{B}_3 \diagdown \widetilde{D}\right) \dot{\cup} \rho\widetilde{D} = \mathbb{B}_3 \diagdown \{0\}.$$

REMARK 12. *The arguments above show that $\mathbb{S}^2$ can be duplicated using 8 pieces, and that $\mathbb{B}_3$ can be duplicated using 16 pieces. More refined arguments show that 4 pieces suffice for $\mathbb{S}^2$, and that 5 pieces suffice for $\mathbb{B}_3$. These latter results are optimal.*

Fourth step: The next result shows that if we declare $A \preceq_G B$ when A is G-equidecomposable to a *subset* of B, then the relation $\preceq_G$ is a partial ordering of the $\sim_G$ equivalence classes in $\mathcal{P}(X)$.

THEOREM 28. *(Banach-Schröder-Bernstein) Suppose that a group G acts on a set X. If $A, B \in \mathcal{P}(X)$ satisfy both $A \preceq_G B$ and $B \preceq_G A$, then $A \sim_G B$.*

Proof: We have the following two properties of the relation $\sim_G$:

- If $A \sim_G B$, then there is a bijection $g : A \to B$ such that

$$C \sim_G g(C) \text{ whenever } C \subset A. \tag{2.2}$$

- If $A_1 \cap A_2 = \phi = B_1 \cap B_2$ and $A_i \sim_G B_i$ for $i = 1, 2$ then $A_1 \cup A_2 \sim_G B_1 \cup B_2$.

By hypothesis, $A \sim_G B_1$ and $A_1 \sim_G B$ for some $B_1 \subset B$ and $A_1 \subset A$. By the first property, there are bijections $f : A \to B_1$ and $g : A_1 \to B$ satisfying $C \sim_G f(C)$ and $D \sim_G g(D)$ whenever $C \subset A$ and $D \subset A_1$. Let $C_0 = A \diagdown A_1$ and inductively $C_{n+1} = g^{-1} f(C_n)$ for $n \geq 0$. With $C = \dot{\cup}_{n=0}^{\infty} C_n$ we have

$$g(A \diagdown C) = B \diagdown f(C)$$

and then $A \diagdown C \sim_G B \diagdown f(C)$ by (2.2). But we also have $C \sim_G f(C)$ by (2.2) and the second property now yields

$$A = (A \diagdown C) \dot{\cup} C \sim_G (B \diagdown f(C)) \dot{\cup} f(C) = B.$$

COROLLARY 12. *A subset E of X is G-paradoxical if and only if there are disjoint sets $A, B \subset E$ with $A \dot{\cup} B = E$ and $A \sim_G E \sim_G B$.*

THEOREM 29. *(strong form of the Banach-Tarski paradox) If A and B are any two bounded subsets of $\mathbb{R}^3$, each with nonempty interior, then A and B are G_3-equidecomposable.*

Proof: It suffices to show that $A \preceq_{G_3} B$, since interchanging A and B yields $B \preceq_{G_3} A$, and the Banach-Schröder-Bernstein theorem then shows that $A \sim_{G_3} B$. So choose solid balls K and L such that $A \subset K$ and $L \subset B$, and let n be large enough that K can be covered by n copies of L. Use the Banach-Tarski paradox to create a union S of n pairwise disjoint copies of L, and then cover K by a union of *translates* of these copies so that $K \preceq_{G_3} S$. It follows that

$$A \subset K \preceq_{G_3} S \sim_{G_3} L \subset B,$$

and so $A \preceq_{G_3} B$.

CHAPTER 6

Abstract integration and the Riesz representation theorem

The properties of Lebesgue measure, as given in Theorem 21, are easily extended to a quite general setting of *measure spaces*, where a theory of integration can then be established that includes the analogues of the monotone convergence theorem, Fatou's lemma and the dominated convergence theorem. It turns out to be fruitful to abandon the completeness property (4) in Theorem 21 for the abstract setting, and to include it as separate feature. The resulting abstract theory of integration is one of the most powerful tools in analysis and we will give several applications of it in the sequel. Fortunately, this abstract theory follows very closely the theory of Lebesgue integration that was developed in the previous chapter, which permits us to proceed relatively quickly here.

1. Abstract integration

Let X be a set and suppose that $\mathcal{A} \subset \mathcal{P}(X)$ is a σ-algebra of subsets of X, i.e. $\mathcal{A}$ contains the empty set, and is closed under complementation and countable unions:

(1) $\phi \in \mathcal{A}$,
(2) $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$,
(3) $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ whenever $A_n \in \mathcal{A}$ for all n.

The pair $(X, \mathcal{A})$ is called a *measurable space* and $\mathcal{A}$ is called a σ-algebra on X, although one usually abuses notation by referring to just X as the measurable space, despite the fact that without $\mathcal{A}$, the set X has no structure. There are lots of σ-algebras on a set X. In fact, given any fixed collection $\mathcal{F} \subset \mathcal{P}(X)$ of subsets of X, there is a *smallest* σ-algebra on X containing $\mathcal{F}$.

Lemma 17. *Given $\mathcal{F} \subset \mathcal{P}(X)$, there is a unique σ-algebra $\mathcal{A}_{\mathcal{F}}$ on X such that*

(1) $\mathcal{F} \subset \mathcal{A}_{\mathcal{F}}$,
(2) *if $\mathcal{A}$ is any σ-algebra on X with $\mathcal{F} \subset \mathcal{A}$, then $\mathcal{A}_{\mathcal{F}} \subset \mathcal{A}$.*

Proof. The power set $\mathcal{P}(X)$ is a σ-algebra on X that contains $\mathcal{F}$. Thus the set

$$\mathcal{A}_{\mathcal{F}} \equiv \bigcap \{\mathcal{A} : \mathcal{A} \text{ is a } \sigma\text{-algebra on } X \text{ with } \mathcal{F} \subset \mathcal{A}\}$$

is nonempty. It is easily verified that $\mathcal{A}_{\mathcal{F}}$ is a σ-algebra on X that contains $\mathcal{F}$, and it is then clear that $\mathcal{A}_{\mathcal{F}}$ is the smallest such. This completes the proof of the lemma. □

A map $\mu : \mathcal{A} \to [0, \infty]$ is called a *positive measure* on $\mathcal{A}$ if it is *countably additive*, and *nondegenerate* in the sense that not every set has infinite measure:

(1) $\dot{\bigcup}_{n=1}^{\infty} E_n \in \mathcal{A}$ and $\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$ whenever $\{E_n\}_{n=1}^{\infty}$ is a pairwise disjoint sequence of sets in $\mathcal{A}$,
(2) there exists $A \in \mathcal{A}$ with $\mu(A) < \infty$.

The triple $(X, \mathcal{A}, \mu)$ is called a *measure space*. Again, one usually abuses notation and refers to such a set functional μ as a positive measure on X, and often as just a measure on X. Note that $\mu(\emptyset) = 0$ is a consequence of properties (1) and (2) since

$$\infty > \mu(A) = \mu\left(A \dot{\cup} \emptyset \dot{\cup} \emptyset \dot{\cup} ...\right) = \mu(A) + \mu(\emptyset) + \mu(\emptyset) + ...$$

and $\mu(A)$ can be cancelled from both sides since $\mu(A) < \infty$. We say that μ is a *complete* measure on X or $\mathcal{A}$ if *all* subsets of sets of μ-measure zero lie in $\mathcal{A}$ and have zero measure, i.e.

(1) $E \in \mathcal{A}$ and $\mu(E) = 0$ whenever $E \subset F$ and $F \in \mathcal{A}$ with $\mu(F) = 0$.

EXAMPLE 4. *We give four examples of measures.*

(1) *Lebesgue measure on the real line* $\mathbb{R}$ *is an example of a complete measure.*
(2) *A simpler example is* counting *measure* $\nu : \mathcal{P}(X) \to [0, \infty]$ *defined on the power set* $\mathcal{P}(X)$ *of a set* X *by*

$$\nu(E) = \begin{cases} \#E & \text{if} \quad E \text{ is finite} \\ \infty & \text{if} \quad E \text{ is infinite} \end{cases}.$$

(3) *Simpler still is the Dirac unit mass measure* $\delta_x : \mathcal{P}(X) \to \{0, 1\}$ *at a point* x *in a set* X *defined by*

$$\delta_x(E) = \begin{cases} 1 & \text{if} \quad x \in E \\ 0 & \text{if} \quad x \notin E \end{cases}.$$

(4) *A very interesting example, and one which often arises as a counterexample to reasonable conjectures in abstract measure theory, uses the well-ordered set* X *that has* ω_1 *as a last element, and with the property that every predessor of* ω_1 *has at most countably many predessors. Recall that an ordered set* $(X, \prec)$ *is* well-ordered *if* $\prec$ *is a linear order on* X *such that every nonempty subset of* X *has a* least *element. The axiom of choice is equivalent to the assertion that* every *set can be well-ordered. To construct* X*, let* Y *be any uncountable well-ordered set and let* ω_1 *be the least element having uncountably many predessors -* ω_1 *is uniquely determined up to order isomorphism and is called the* first uncountable ordinal*.*

Now for $\alpha \in X$*, let* P_α *and* S_α *be the predessor and successor sets of* α *given by*

$$\begin{aligned} P_\alpha &\equiv \{\beta \in X : \beta \prec \alpha\}, \\ S_\alpha &\equiv \{\beta \in X : \alpha \prec \beta\}. \end{aligned}$$

Define a topology τ *on* X *by declaring that* $G \subset X$ *belongs to* τ *if* G *is either a predessor set* P_α*, a successor set* S_β*, an open segment* $P_\alpha \cap S_\beta \equiv (\beta, \alpha)$*, or an arbitrary union of predessors, successors and segments. Then the topological space* (X, τ) *is Hausdorff (meaning that every pair of distinct points in* X *can be separated by disjoint open sets in* X*) and compact.*

To see that X is compact, observe that every collection of closed subsets $\{F_i\}_{i\in I}$ with the finite intersection property has nonempty *intersection, $\bigcap_{i\in I} F_i \neq \emptyset$, because every nonempty subset of X has a least element. Indeed, if it were the case that $\bigcap_{i\in I} F_i = \emptyset$, then there is an infinite sequence $\{F_{i_n}\}_{n=1}^{\infty}$ of these closed sets, such that the* least upper bounds α_n *of the sets $\bigcap_{k=1}^{n} F_{i_k}$ form an infinite* strictly decreasing *sequence $\{\alpha_n\}_{n=1}^{\infty}$ in X, contradicting the existence of a least element in $\{\alpha_n\}_{n=1}^{\infty}$. To see this, choose F_{i_1} arbitarily. Then $\alpha_1 \equiv lub\,(F_{i_1})$ exists and lies in F_{i_1}. In fact, the set of upper bounds of* any *set F is nonempty (ω_1 is an upper bound), and so has a least element α because X is well-ordered. Every open set containing α must contain a segment (β, γ) with $\beta < \alpha < \gamma$ (or a predessor or successor set containing α - we leave these cases to the reader), and since β cannot be an upper bound for F, there is an element of F in the segment $(\beta, \alpha]$. If F is closed it thus follows that $\alpha \in F$. Next, we note that there is F_{i_2} such that $\alpha_1 \notin F_{i_1} \cap F_{i_2}$ (otherwise $\bigcap_{i\in I} F_i \neq \emptyset$), and since $F_{i_1} \cap F_{i_2}$ is closed and nonempty, we have*

$$\alpha_2 \equiv lub\,(F_{i_1} \cap F_{i_2}) < \alpha_1.$$

We can continue in this manner to construct a sequence of sets $\{F_{i_n}\}_{n=1}^{\infty}$ such that the points $\alpha_n \equiv lub\,(\bigcap_{k=1}^{n} F_{i_k})$ are strictly decreasing.

Now define

$$\mathcal{A} \equiv \left\{ \begin{array}{r} E \in \mathcal{P}(X) : \text{either } E \cup \{\omega_1\} \text{ or } E^c \cup \{\omega_1\} \\ \text{contains an uncountable compact set} \end{array} \right\},$$

and define a set functional $\lambda : \mathcal{A} \to \{0,1\}$ by

$$\lambda(E) = \left\{ \begin{array}{lll} 1 & \text{if} & E \cup \{\omega_1\} \text{ contains an uncountable compact set} \\ 0 & \text{if} & E^c \cup \{\omega_1\} \text{ contains an uncountable compact set} \end{array} \right.,$$

for $E \in \mathcal{A}$.

EXERCISE 2. *With regard to Example (4) above, prove that $\mathcal{A}$ is a σ-algebra on X containing the open sets, and that λ is a positive measure on $\mathcal{A}$.* Hint*: Show that every countable intersection of uncountable compact subsets of X is uncountable. Recall that compact sets are closed in a Hausdorff space such as X.*

1.1. Measurable functions. It is convenient to initially define the notion of a *measurable function* for $f : X \to Y$ where X is a measure space and Y is a general topological space. Recall that $\tau \subset \mathcal{P}(Y)$ is a *topology* on Y if it contains the empty set, the whole set Y, and is closed under finite intersections and arbitrary unions:

(1) $\emptyset, X \in \tau$,
(2) $G_1 \cap G_2 \cap ... \cap G_n \in \tau$ whenever $G_i \in \tau$ for $1 \leq i \leq n < \infty$,
(3) $\bigcup_{\alpha\in A} G_\alpha \in \tau$ whenever $G_\alpha \in \tau$ for all $\alpha \in A$ (here A is an arbitrary index set).

The pair (Y, τ) is called a topological space, and the sets in τ are called the open sets in Y. As usual, we often abuse notation and refer to just the set Y as the topological space, with the underlying topology being understood.

DEFINITION 14. *Let $(X, \mathcal{A})$ be a measurable space and let (Y, τ) be a topological space. A function*

$$f : X \to Y$$

is said to be measurable (more precisely $\mathcal{A}$-measurable) if

$$f^{-1}(G) \in \mathcal{A} \text{ for all } G \in \tau.$$

Note the similarity to the definition of a *continuous* function $f : X \to Y$ in the case that (X, σ) is a topological space: f is continuous if $f^{-1}(G) \in \sigma$ for all $G \in \tau$. We have already seen in Example 3 that the composition of a continuous function followed by a Lebesgue measurable function need not be measurable. On the other hand the composition of a measurable function followed by a continuous function is *always* measurable, even in this abstract setting.

PROPOSITION 5. *Suppose that $(X, \mathcal{A})$ is a measurable space and that (Y, τ) and (Z, μ) are topological spaces. If $f : X \to Y$ is measurable and $g : Y \to Z$ is continuous, then the composition $g \circ f : X \to Z$ is measurable.*

PROOF. If $H \in \mu$ is open in Z, then $G \equiv g^{-1}(H) \in \sigma$ is open in Y and so the measurability of f gives

$$(g \circ f)^{-1}(H) = f^{-1}\left(g^{-1}(H)\right) = f^{-1}(G) \in \mathcal{A}$$

for all $H \in \mu$. This verifies the definition that $g \circ f : X \to Z$ is measurable. □

We now consider the possibility that X is simultaneously a measurable space and a topological space, i.e. there is a σ-algebra $\mathcal{A}$ on X as well as a topology τ on X. If $\tau \subset \mathcal{A}$, then every continuous function $f : X \to Y$ is also measurable.

LEMMA 18. *Suppose that $\mathcal{A}$ is a σ-algebra on X and τ is a topology on X with $\tau \subset \mathcal{A}$. If Y is any topological space, then every continuous function $f : X \to Y$ is also measurable.*

PROOF. If G is open in Y, then $f^{-1}(G) \in \tau \subset \mathcal{A}$. □

If (X, τ) is a topological space, then Lemma 17 shows that there is a smallest σ-algebra $\mathcal{B}_\tau$ on X that contains the topology τ. This important σ-algebra $\mathcal{B}_\tau$ is called the *Borel* σ-algebra on the topological space (X, τ), and the sets E in $\mathcal{B}_\tau$ are called *Borel sets*. A function $f : X \to Y$ that is measurable with respect to the Borel σ-algebra on X is said to be a Borel function on X. The previous lemma shows that continuous functions are always Borel measurable, but there is an important property that Borel functions have that is *not* shared by measurable functions in general.

PROPOSITION 6. *Suppose that $(X, \mathcal{A})$ is a measurable space and that (Y, τ) and (Z, μ) are topological spaces. If $f : X \to Y$ is measurable and $g : Y \to Z$ is Borel measurable, then the composition $g \circ f : X \to Z$ is measurable.*

PROOF. Consider the collection of subsets of Y defined by

$$\mathcal{C} \equiv \left\{B \in \mathcal{P}(Y) : f^{-1}(B) \in \mathcal{A}\right\}.$$

It is a simple exercise to verify that $\mathcal{C}$ is a σ-algebra on Y (no properties other than $\mathcal{A}$ is a σ-algebra and f is a function are needed for this). Indeed, the following three properties hold since $\mathcal{A}$ is a σ-algebra;

$$\begin{aligned}
f^{-1}(\emptyset) &= \emptyset \in \mathcal{A}, \\
f^{-1}(B^c) &= \left[f^{-1}(B)\right]^c \in \mathcal{A}, \qquad \text{if } B \in \mathcal{C}, \\
f^{-1}\left(\bigcup_{k=1}^{\infty} B_k\right) &= \bigcup_{k=1}^{\infty} f^{-1}(B_k) \in \mathcal{A}, \qquad \text{if } B_k \in \mathcal{C},
\end{aligned}$$

and they show by definition of $\mathcal{C}$ that

$$\begin{aligned}\emptyset &\in \mathcal{C},\\ B^c &\in \mathcal{C} \text{ when } B \in \mathcal{C},\\ \bigcup_{k=1}^{\infty} B_k &\in \mathcal{C} \text{ when } B_k \in \mathcal{C}.\end{aligned}$$

Moreover, the measurability of f shows that $\mathcal{C}$ contains τ, the open sets in Y. Thus by Lemma 17, $\mathcal{C}$ contains the Borel σ-algebra $\mathcal{B}_\tau$ on Y.

Now if $H \in \mu$ is open in Z, the Borel measurability of g shows that

$$g^{-1}(H) \in \mathcal{B}_\tau \subset \mathcal{C},$$

which gives

$$(g \circ f)^{-1}(H) = f^{-1}\left(g^{-1}(H)\right) \in \mathcal{A}$$

for every $H \in \mu$ by the definition of $g^{-1}(H) \in \mathcal{C}$. □

REMARK 13. *For future reference we isolate one of the facts proved above: if $\mathcal{A}$ is a σ-algebra on a set X, and if $f : X \to Y$ is* any *function whatsoever, then the set*

$$\mathcal{C} \equiv \left\{B \in \mathcal{P}(Y) : f^{-1}(B) \in \mathcal{A}\right\}$$

is a σ-algebra on Y. Thus σ-algebras can be pushed forward *by arbitrary functions.*

1.1.1. *Product spaces.* Given two topological spaces (Y_1, τ_1) and (Y_2, τ_2), we define the *product topology* $\tau_1 \times \tau_2$ on the product space $Y_1 \times Y_2$ to consist of arbitrary unions of *open rectangles* $G_1 \times G_2$ where $G_i \in \tau_i$ for $i = 1, 2$. It is easy to see that $\tau_1 \times \tau_2$ is a topology - it is closed under finite intersections since the intersection of two open rectangles is again an open rectangle. Let (X, σ) be another topological space. It is an easy exercise to show that if

$$f : X \to Y_1 \times Y_2, \qquad f(x) = (f_1(x), f_2(x)) \in Y_1 \times Y_2 \text{ for } x \in X,$$

then f is continuous *if and only if* $f_i : X \to Y_i$ is continuous for both $i = 1$ and $i = 2$. The same sort of phenomenon holds for measurability if the spaces Y_1 and Y_2 each have a countable base. Recall that a topological space (Y, τ) has a *countable base* $\{G_n\}_{n=1}^{\infty}$ if each G_n is open and if for every point x contained in an open set G there is $n \in \mathbb{N}$ such that $x \in G_n \subset G$. For example, Euclidean space $\mathbb{R}^m$ has a countable base, namely the collection of all open balls with rational radii having centers with rational coordinates.

LEMMA 19. *Suppose that $(X, \mathcal{A})$ is a measurable space, and that (Y_1, τ_1) and (Y_2, τ_2) are topological spaces with countable bases. Then*

$$f = (f_1, f_2) : X \to Y_1 \times Y_2$$

is measurable if and only if *$f_i : X \to Y_i$ is measurable for both $i = 1$ and $i = 2$.*

PROOF. Suppose first that f is measurable. Since the projection map $\pi_i : Y_1 \times Y_2 \to Y_i$ is continuous, Proposition 5 shows that $f_i = \pi_i \circ f$ is measurable for $i = 1, 2$.

Now suppose that both f_1 and f_2 are measurable. Then if $R = G_1 \times G_2$ is an open rectangle,

$$f^{-1}(R) = f^{-1}(G_1 \times G_2) = f_1^{-1}(G_1) \cap f_2^{-1}(G_2) \in \mathcal{A}. \tag{1.1}$$

If $J=\bigcup_{k=1}^{\infty} R_k$ is a *countable* union of open rectangles R_k, we have

$$f^{-1}(J)=\bigcup_{k=1}^{\infty} f^{-1}(R_k)\in\mathcal{A}.$$

Finally, it is easy to see that every open set J in $Y_1\times Y_2$ is a countable union of open rectangles because of our assumption that Y_i has a countable base for $i=1$ and $i=2$. Indeed, if $\mathcal{B}_i$ is a countable base for Y_i, then

$$\mathcal{B}_1\times\mathcal{B}_2\equiv\{G_1\times G_2:G_i\in\mathcal{B}_i \text{ for } 1=1,2\}$$

is a countable base for $Y_1\times Y_2$. Then if J is open,

$$J=\bigcup\{G:G\in\mathcal{B}_1\times\mathcal{B}_2 \text{ and } G\subset J\},$$

and the latter union is clearly at most countable. This completes the proof that f is measurable. □

COROLLARY 13. *Let $(X,\mathcal{A})$ be a measurable space and $n\geq 2$. Then*

(1) *$f:X\to\mathbb{R}^n$ is measurable* if and only if *each component function $f_i:X\to\mathbb{R}$ in $f(x)=(f_1(x),...,f_n(x))$ is measurable, $1\leq i\leq n$, and*
(2) *if $f,g:X\to\mathbb{R}^n$ are both measurable, then so are $f+g:X\to\mathbb{R}^n$ and $f\cdot g:X\to\mathbb{R}$.*

PROOF. Assertion (1) follows by induction from Lemma 19. Now define $F:X\to\mathbb{R}^n\times\mathbb{R}^n$ by $F(x)\equiv(f(x),g(x))$ for $x\in X$. Then the measurability of f and g implies that of F by Lemma 19. If $\varphi:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}^n$ is defined by $\varphi(u,v)=u+v$, then the continuity of φ and Proposition 5 imply the measurability of $(\varphi\circ F)(x)=f(x)+g(x)=(f+g)(x)$, $x\in X$. Similarly, if $\psi:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ is defined by $\varphi(u,v)=u\cdot v$, then the continuity of ψ and Proposition 5 imply the measurability of $(\psi\circ F)(x)=f(x)\cdot g(x)=(f\cdot g)(x)$, $x\in X$. □

The following lemma is proved exactly as in the case of Lebesgue measure on the real line treated above.

LEMMA 20. *Let $(X,\mathcal{A})$ be a measurable space. Suppose that $f,f_n,g:X\to[-\infty,\infty]$ for $n\in\mathbb{N}$.*

(1) *If f is finite-valued, then f is measurable* if and only if *$f^{-1}(G)\in\mathcal{A}$ for all open sets $G\subset\mathbb{R}$* if and only if *$f^{-1}(F)\in\mathcal{A}$ for all closed sets $F\subset\mathbb{R}$.*
(2) *If f is finite-valued and continuous, then f is measurable.*
(3) *If f is finite-valued and measurable and $\Phi:\mathbb{R}\to\mathbb{R}$ is continuous, then $\Phi\circ f$ is measurable.*
(4) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions, then the following functions are all measurable:*

$$\sup_n f_n(x),\quad \inf_n f_n(x),...\lim\sup_{n\to\infty} f_n(x),\quad \lim\inf_{n\to\infty} f_n(x).$$

(5) *If $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions and $f(x)=\lim_{n\to\infty}f_n(x)$, then f is measurable.*
(6) *If f is measurable, so is f^n for $n\in\mathbb{N}$.*
(7) *If f and g are finite-valued and measurable, then so are $f+g$ and fg.*

1.2. Simple, nonnegative and integrable functions. We now proceed almost exactly as we did in the case of Lebesgue measure on the real line $\mathbb{R}$. We will be brief and omit all proofs here as they are virtually verbatim the same as the proofs we gave for Lebesgue measure.

Let $(X, \mathcal{A}, \mu)$ be a measure space. A function $\varphi : X \to \mathbb{R}$ is a *simple* function if it is measurable and its range is finite. Such functions have the form

$$\varphi = \sum_{k=1}^{N} \alpha_k \chi_{E_k}, \qquad \alpha_k \in \mathbb{R}, E_k \in \mathcal{A}.$$

PROPOSITION 7. *Let $f : X \to [-\infty, \infty]$ be measurable.*

(1) *If f is nonnegative there is an increasing sequence of nonnegative simple functions $\{\varphi_k\}_{k=1}^{\infty}$ that converges pointwise and monotonically to f:*

$$\varphi_k(x) \leq \varphi_{k+1}(x) \text{ and } \lim_{k\to\infty} \varphi_k(x) = f(x), \qquad \text{for all } x \in X.$$

(2) *There is a sequence of simple functions $\{\varphi_k\}_{k=1}^{\infty}$ satisfying*

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)| \text{ and } \lim_{k\to\infty} \varphi_k(x) = f(x), \qquad \text{for all } x \in X.$$

If φ is a simple function, then φ has a unique canonical representation

$$\varphi = \sum_{k=1}^{N} \alpha_k \chi_{E_k},$$

where the real constants α_k are distinct and nonzero, and the measurable sets E_k are pairwise disjoint. We define the integral of φ by

$$\int \varphi d\mu = \sum_{k=1}^{N} \alpha_k |E_k|_{\mu},$$

where we are using the notation $|E_k|_{\mu} = \mu(E)$ for $E \in \mathcal{A}$. If $E \in \mathcal{A}$ and φ is a simple function, then so is $\chi_E \varphi$, and we define

$$\int_E \varphi d\mu = \int (\chi_E \varphi)\, d\mu.$$

LEMMA 21. *Suppose that φ and ψ are simple functions and that $E, F \in \mathcal{A}$.*

(1) *If $\varphi = \sum_{k=1}^{M} \beta_k \chi_{F_k}$ (not necessarily the canonical representation), then*

$$\int \varphi d\mu = \sum_{k=1}^{M} \beta_k |F_k|_{\mu}.$$

(2) $\int (a\varphi + b\psi)\, d\mu = a \int \varphi d\mu + b \int \psi d\mu$ *for $a, b \in \mathbb{C}$,*
(3) $\int_{E\cup F} \varphi d\mu = \int_E \varphi d\mu + \int_F \varphi$ *if $E \cap F = \phi$,*
(4) $\int \varphi d\mu \leq \int \psi d\mu$ *if $\varphi \leq \psi$,*
(5) $\left|\int \varphi d\mu\right| \leq \int |\varphi|\, d\mu$.

For $f : X \to [0, \infty]$ measurable we define

$$\int f d\mu = \sup \left\{ \int \varphi d\mu : 0 \leq \varphi \leq f \text{ and } \varphi \text{ is simple} \right\}.$$

We say that f is *integrable* if $\int f d\mu < \infty$. For E measurable define

$$\int_E f d\mu = \int (\chi_E f)\, d\mu.$$

LEMMA 22. *Suppose that $f, g : X \to [0, \infty]$ are nonnegative measurable functions and that $E, F \in \mathcal{A}$.*

(1) $\int (af + bg)\, d\mu = a \int f d\mu + b \int g d\mu$ *for* $a, b \in (0, \infty)$,
(2) $\int_{E \cup F} f d\mu = \int_E f d\mu + \int_F f d\mu$ *if* $E \cap F = \phi$,
(3) $\int f d\mu \leq \int g d\mu$ *if* $0 \leq f \leq g$,
(4) *If* $\int f d\mu < \infty$, *then* $f(x) < \infty$ *for a.e.* x,
(5) *If* $\int f d\mu = 0$, *then* $f(x) = 0$ *for a.e.* x.

THEOREM 30. *(Monotone Convergence Theorem) Suppose that $\{f_n\}_{n=1}^{\infty}$ is an* increasing *sequence of nonnegative measurable functions, i.e. $f_n(x) \leq f_{n+1}(x)$, and let*

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

Then f is nonegative and measurable and

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

COROLLARY 14. *Suppose that $a_k(x) \geq 0$ is measurable for $k \geq 1$. Then*

$$\int \sum_{k=1}^{\infty} a_k d\mu = \sum_{k=1}^{\infty} \int a_k d\mu.$$

LEMMA 23. *(Fatou's Lemma) If $\{f_n\}_{n=1}^{\infty}$ is a sequence of* nonnegative *measurable functions, then*

$$\int \lim \inf_{n \to \infty} f_n d\mu \leq \lim \inf_{n \to \infty} \int f_n d\mu.$$

If $f : X \to [-\infty, \infty]$ is measurable, define

$$\int f d\mu = \int f^+ d\mu - \int f_- d\mu,$$

provided not both $\int f^+ d\mu$ and $\int f_- d\mu$ are infinite. We say that such an f is integrable if

$$\int |f|\, d\mu = \int \left(f^+ + f_-\right) d\mu = \int f^+ d\mu + \int f_- d\mu < \infty.$$

LEMMA 24. *Suppose that f, g are integrable and that $E, F \in \mathcal{A}$.*

(1) $\int (af + bg)\, d\mu = a \int f d\mu + b \int g d\mu$ *for* $a, b \in \mathbb{R}$,
(2) $\int_{E \cup F} f d\mu = \int_E f d\mu + \int_F f$ *if* $E \cap F = \phi$,
(3) $\int f d\mu \leq \int g d\mu$ *if* $f \leq g$,
(4) $\left|\int f d\mu\right| \leq \int |f|\, d\mu$.

We say that a property $P(x)$ holds $\mu - a.e.$ $x \in X$ if the set of x for which $P(x)$ fails has μ-measure zero.

THEOREM 31. *(Dominated Convergence Theorem) Let g be a nonnegative integrable function. Suppose that $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions satisfying*

$$\lim_{n\to\infty} f_n(x) = f(x), \qquad \mu - a.e.\ x \in X,$$

and

$$|f_n(x)| \le g(x), \qquad \mu - a.e.\ x \in X.$$

Then

$$\lim_{n\to\infty} \int |f - f_n|\, d\mu = 0,$$

and hence

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

Finally, if $f(x) = u(x) + iv(x)$ is complex-valued where $u(x)$ and $v(x)$ are real-valued measurable functions such that

$$\int |f|\, d\mu = \int \sqrt{u^2+v^2} d\mu < \infty,$$

then we define

$$\int f d\mu = \int u d\mu + i \int v d\mu.$$

The usual properties of linearity, additivity, monotonicity and the triangle inequality all hold for this definition as well.

2. The Riesz representation theorem

Suppose we have a measure space $(X, \mathcal{A}, \mu)$ that is also a topological space (X, τ) with topology $\tau \subset \mathcal{A}$. Then every continuous function $f : X \to \mathbb{C}$ is measurable. If in addition the measure μ is *locally finite*, i.e.

$$\mu(K) < \infty \text{ for all compact sets } K \subset X,$$

and if the space X is compact, or more generally just if f has compact support, then f is integrable and the integral $\int f d\mu$ is a complex number. Now the set $C_c(X)$ of continuous complex-valued functions on X with compact support is clearly a complex vector space under pointwise addition and scalar multiplication of functions. The map

$$\Lambda_\mu : C_c(X) \to \mathbb{C}, \text{ given by } \Lambda_\mu f = \int f d\mu,$$

is a *linear functional* on the vector space $C_c(X)$. Moreover it has a special property due to the positivity of the measure μ, namely that Λ_μ is a *positive* linear functional:

$$\Lambda_\mu f \ge 0 \text{ whenever } f \in C_c(X) \text{ satisfies } f(x) \ge 0 \text{ for all } x \in X.$$

Remarkable fact: For many topological spaces (X, τ), every positive linear functional Λ on $C_c(X)$ is equal to Λ_μ for some positive locally finite Borel measure μ on X.

The condition we will impose on the space X in order to force this remarkable fact is that X be locally compact and Hausdorff. A topological space (X, τ) is *locally compact* if X has a base of compact sets, i.e. for every $x \in G \subset X$ with G open, there is H open with $\overline{H}$ compact and

$$x \in H \subset \overline{H} \subset G \subset X.$$

A topological space (X,τ) is *Hausdorff* if for every pair of distinct points $x, y \in X$ there are open sets G and H such that

$$x \in G,\ y \in H \text{ and } G \cap H = \emptyset.$$

The key fact that we use about such spaces, and which connects measures to continuous functions is *Urysohn's Lemma.*

2.1. Urysohn's Lemma.

LEMMA 25. *(Urysohn) Suppose that X is a locally compact Hausdorff space and that $K \subset V \subset X$ where K is compact and V is open. Then there is a continuous function with compact support $f \in C_c(X)$ such that*

$$\chi_K(x) \leq f(x) \leq \chi_V(x), \qquad x \in X. \tag{2.1}$$

In particular $f = 1$ on K and $f = 0$ outside V.

The conclusion of Urysohn's Lemma can be viewed as a strong form of the Hausdorff property. It says that if K is a compact set and F is a closed set, then K and F can be 'separated' by a continuous function that is 0 on F and 1 on K. In particular, if singletons are closed in X, then given x and y distinct points in X, we can take $K = \{x\}$, $F = \{y\}$, $G = \left\{f > \frac{1}{2}\right\}$ and $H = \left\{f < \frac{1}{2}\right\}$ to obtain the Hausdorff property $x \in G$, $y \in H$ and $G \cap H = \emptyset$.

Proof of Urysohn's Lemma: We give the proof in three steps.

Step 1: We first show that we can squeeze an open set U with compact closure $\overline{U}$ between K and V as follows:

$$K \subset U \subset \overline{U} \subset V. \tag{2.2}$$

Here is how to construct such a set U. For each $p \in K$ we use the fact that X is *locally compact* to choose an open set O_p containing p and such that $\overline{O_p}$ is compact. Since K is compact there is a finite collection $\{O_{p_n}\}_{n=1}^N$ of these open sets that cover K. Then $O \equiv \bigcup_{n=1}^N O_{p_n}$ is open, contains K and $\overline{O} = \bigcup_{n=1}^N \overline{O_{p_n}}$ is compact. In the special case that $V = X$ we can take $U = O$. Otherwise, $F = X \setminus V$ is closed and nonempty.

Now we use the *Hausdorff* property of X to obtain that for every $x \in K$ and $y \in F$, there are open sets $G_{(x,y)}$ and $H_{(x,y)}$ with

$$x \in G_{(x,y)},\ y \in H_{(x,y)} \text{ and } G_{(x,y)} \cap H_{(x,y)} = \emptyset.$$

Momentarily fix $y \in F$. Since K is compact there is a finite subcollection $\left\{G_{(x_m,y)}\right\}_{m=1}^M$ that covers X. Then the open sets

$$G^y \equiv \bigcup_{m=1}^M G_{(x_m,y)} \text{ and } H^y = \bigcap_{m=1}^M H_{(x_m,y)}$$

separate K and y in the sense that G^y and H^y are disjoint open sets that contain K and y respectively. Thus $y \notin \overline{G^y}$ and we see that the collection of sets $\left\{F \cap \overline{O} \cap \overline{G^y}\right\}_{y \in F}$ satisfies

$$\bigcap_{y \in F} \left(F \cap \overline{O} \cap \overline{G^y}\right) = \emptyset.$$

Since the sets $F \cap \overline{O} \cap \overline{G^y}$ are compact ($\overline{O}$ is compact and F and $\overline{G^y}$ are closed), the *finite intersection property* shows that there is a finite subcollection $\left\{F \cap \overline{O} \cap \overline{G^{y_\ell}}\right\}_{\ell=1}^{L}$ satisfying

$$\bigcap_{\ell=1}^{L} \left(F \cap \overline{O} \cap \overline{G^{y_\ell}}\right) = \emptyset.$$

Then the set

$$U \equiv \bigcap_{\ell=1}^{L} \left(O \cap G^{y_\ell}\right)$$

is open with compact closure $\overline{U} = \bigcap_{\ell=1}^{L} \left(\overline{O} \cap \overline{G^{y_\ell}}\right)$, and of course

$$\overline{U} \subset X \setminus F = V.$$

Step 2: We now iterate the squeezing process as follows. First rewrite (2.2) with U_0 in place of U:

$$K \subset U_0 \subset \overline{U_0} \subset V.$$

Then apply (2.2) to the pair of sets $K \subset U_0$ where K is compact and U_0 is open to obtain an open set U_1 with compact closure satisfying

$$K \subset U_1 \subset \overline{U_1} \subset U_0 \subset \overline{U_0} \subset V.$$

Next, apply (2.2) to the pair of sets $\overline{U_1} \subset U_0$ where $\overline{U_1}$ is compact and U_0 is open to obtain an open set $U_{\frac{1}{2}}$ with compact closure satisfying

$$K \subset U_1 \subset \overline{U_1} \subset U_{\frac{1}{2}} \subset \overline{U_{\frac{1}{2}}} \subset U_0 \subset \overline{U_0} \subset V.$$

We continue with

$$K \subset U_1 \subset \overline{U_1} \subset U_{\frac{3}{4}} \subset \overline{U_{\frac{3}{4}}} \subset U_{\frac{1}{2}} \subset \overline{U_{\frac{1}{2}}} \subset U_{\frac{1}{4}} \subset \overline{U_{\frac{1}{4}}} \subset U_0 \subset \overline{U_0} \subset V,$$

and then

$$\begin{aligned} K &\subset U_1 \subset \overline{U_1} \subset U_{\frac{7}{8}} \subset \overline{U_{\frac{7}{8}}} \subset U_{\frac{3}{4}} \subset \overline{U_{\frac{3}{4}}} \\ &\subset U_{\frac{5}{8}} \subset \overline{U_{\frac{5}{8}}} \subset U_{\frac{1}{2}} \subset \overline{U_{\frac{1}{2}}} \subset U_{\frac{3}{8}} \subset \overline{U_{\frac{3}{8}}} \\ &\subset U_{\frac{1}{4}} \subset \overline{U_{\frac{1}{4}}} \subset U_{\frac{1}{8}} \subset \overline{U_{\frac{1}{8}}} \subset U_0 \subset \overline{U_0} \subset V. \end{aligned}$$

This process can be continued indefinitely and produces a collection of open sets $\{U_r\}_{r \in \mathcal{D}}$ where $\mathcal{D} = \left\{\frac{k}{2^\ell} : k, \ell \in \mathbb{N} \text{ with } \ell \geq 1 \text{ and } 0 \leq k \leq 2^\ell\right\}$ and that satisfies the property

$$K \subset U_r \subset \overline{U_r} \subset U_s \subset \overline{U_s} \subset V \tag{2.3}$$

whenever $r, s \in \mathcal{D}$ with $r > s$.

Step 3: We can now define our candidate for the function $f : X \to [0,1]$ in (2.1). Given $x \in X$ we define

$$f(x) \equiv \sup_{0 \leq r \leq 1} r \chi_{U_r}(x).$$

Then we have

$$\{x \in X : f(x) > \lambda\} = \bigcup_{r > \lambda} U_r \text{ is } \textit{open} \text{ for all } 0 \leq \lambda \leq 1.$$

We similarly have that the function g defined by

$$g(x) \equiv \inf_{0 \le s \le 1} s\chi_{\left(\overline{U_s}\right)^c}(x),$$

satisfies

$$\{x \in X : g(x) < \lambda\} = \bigcup_{s<\lambda} \left(\overline{U_s}\right)^c \text{ is } \textit{open} \text{ for all } 0 \le \lambda \le 1.$$

If we can show that $f = g$, it will then follow that f is continuous since

$$f^{-1}((a,b)) = \{f > a\} \cap \{g < b\}$$

will be open for all open intervals (a,b) with $0 \le a < b \le 1$, and this is enough to establish the continuity of $f : X \to [0,1]$. Now it suffices to show that $f(x) = g(x)$ for all $x \in U_0 \setminus \overline{U_1}$ since both f and g vanish outside U_0, and both are 1 inside $\overline{U_1}$. But if $f(x) > g(x)$ and $x \in U_0 \setminus \overline{U_1}$, then there is $r > s$ such that $x \in U_r$ and $x \in \left(\overline{U_s}\right)^c$, which implies $U_r \supset \overline{U_s}$, contradicting (2.3) which says $\overline{U_r} \subset U_s$. On the other hand, if $f(x) < g(x)$ and $x \in U_0 \setminus \overline{U_1}$, then there are $t, v \in \mathcal{D}$ such that $f(x) < t < v < g(x)$ with $x \notin \overline{U_t}$ and $x \notin (U_v)^c$. Thus $x \in \left(\overline{U_t}\right)^c \cap U_v$ which implies $v < t$, contradicting our assumption that $t < v$. This completes the proof that $f = g$, and hence the proof of Urysohn's Lemma.

Urysohn's Lemma can be thought of as a continuous *unit* function on the compact set K that is *subordinate* to the open set V covering K. A simple algebraic trick permits us to obtain a far more flexible variant, namely a continuous *partition of unity* on the compact set K that is subordinate to a finite open cover $\{V_n\}_{n=1}^N$ of K.

COROLLARY 15. *Suppose that $\{V_n\}_{n=1}^N$ is a finite collection of open subsets of a locally compact Hausdorff space X. If K is a compact subset of X that is covered by $\{V_n\}_{n=1}^N$, then there exist continuous compactly supported functions $\{f_n\}_{n=1}^N \subset C_c(X)$ satisfying*

$$\begin{aligned} \chi_K &\le \sum_{n=1}^N f_n \le \chi_{\bigcup_{n=1}^N V_n}, \\ 0 &\le f_n \le \chi_{V_n}, \qquad 1 \le n \le N. \end{aligned}$$

In particular, $\sum_{n=1}^N f_n = 1$ on K and $f_n = 0$ outside V_n.

Proof: For each $x \in K$ there is $n = n(x)$ such that $x \in V_n$. Since X is locally compact, there is an open set $W_{x,n}$ with compact closure satisfying $x \in W_{x,n} \subset \overline{W_{x,n}} \subset V_{n(x)}$. Then $\left\{W_{x,n(x)}\right\}_{x \in X}$ is an open cover of K, and since K is compact, there is a finite subcover $\{G_\ell\}_{\ell=1}^L$. Now for $1 \le n \le N$ let J_n be the union of all $\overline{G_\ell}$ that are contained in V_n. Then J_n is a finite union of compact sets, so is compact itself. Also, we have $K \subset \bigcup_{n=1}^N J_n$. Indeed, every $y \in K$ lies in some G_ℓ, and G_ℓ equals $W_{x,n(x)}$ for some $x \in X$, and so $y \in J_{n(x)}$ since

$$y \in G_\ell = W_{x,n(x)} \subset \overline{W_{x,n(x)}} \subset V_{n(x)}.$$

Now apply Urysohn's Lemma to the pair $J_n \subset V_n$ to obtain $g_n \in C_c(x)$ such that

$$\chi_{J_n} \le g_n \le \chi_{V_n}, \qquad 1 \le n \le N.$$

Now we use an algebraic trick motivated by the solution to a well known mathematical teaser of P. Halmos.

Mathematical Teaser: A barrel of pickles that is 99% water by weight is opened at sunrise and left out in the sun all day. At sunset it is 98% water and weighs 500 lbs. How much did the barrel weigh at sunrise?

Solution: *Consider the* complement *of the water. Since the percentage of nonwater in the barrel doubles during the day (it goes from 1% nonwater to 2% nonwater), the weight of the barrel and contents must have been cut in half by sunset (the weight of nonwater - the barrel and pickle pulp - remains constant). Thus the barrel started the day at 1000 lbs.*

To apply this principle of *complementation* to our partition of unity problem, we observe that each continuous function $1-g_n$ vanishes on J_n, hence the *product* $\prod_{n=1}^{N}(1-g_n)$ is continuous and vanishes on $\bigcup_{n=1}^{N} J_n$. Thus $f \equiv 1-\prod_{n=1}^{N}(1-g_n)$ is continuous and equals 1 on K and vanishes outside $\bigcup_{n=1}^{N} V_n$. It remains only to write $f=\sum_{n=1}^{N} f_n$ where each f_n is continuous and satisfies $0 \leq f_n \leq \chi_{V_n}$. But this can be achieved by writing

$$\begin{aligned}\prod_{n=1}^{N}(1-g_n) &= -g_N \prod_{n=1}^{N-1}(1-g_n)+\prod_{n=1}^{N-1}(1-g_n) \\ &= -g_N \prod_{n=1}^{N-1}(1-g_n)-g_{N-1}\prod_{n=1}^{N-2}(1-g_n)+\prod_{n=1}^{N-2}(1-g_n) \\ &= -f_N-f_{N-1}-\dots-f_1+1,\end{aligned}$$

where

$$\begin{aligned} f_1 &= g_1, \\ f_2 &= (1-g_1)\,g_2, \\ f_3 &= (1-g_1)(1-g_2)\,g_3, \\ &\vdots \\ f_N &= (1-g_1)(1-g_2)\dots(1-g_{N-1})\,g_N. \end{aligned}$$

Of course we could have simply *begun* by defining f_n as above, and then using induction on n to show that

$$1-\sum_{k=1}^{n} f_k=\prod_{k=1}^{n}(1-g_k), \qquad 1 \leq n \leq N.$$

However, this would have denied us the fun of finding the formulas in the first place. In any event, the case $n=N$ yields

$$\sum_{k=1}^{N} f_k(x)=1-\prod_{k=1}^{N}(1-g_k)=1, \qquad x \in K,$$

since every $x \in K$ lies in J_k for some $1 \leq k \leq N$, and hence $1-g_k(x)=1-1=0$.

Finally $\chi_{J_n} \leq g_n \leq \chi_{V_n}$ and $0 \leq \prod_{k=1}^{n}(1-g_k) \leq 1$ imply

$$0 \leq \left\{\prod_{k=1}^{n}(1-g_k)\right\}\chi_{J_n} \leq \left\{\prod_{k=1}^{n}(1-g_k)\right\} g_n = f_n \leq \chi_{V_n}$$

for all $1 \leq n \leq N$.

2.2. Representing continuous linear functionals. In preparation for stating the Riesz representation theorem, we introduce some regularity terminology that links measure and topology.

DEFINITION 15. *Suppose that μ is a* Borel *measure on a topological space X. We say μ is* outer regular *if*

$$\mu(E) = \inf\{\mu(V) : E \subset V \text{ open}\} \tag{2.4}$$

for all Borel sets E. We say μ is inner regular *if*

$$\mu(E) = \sup\{\mu(K) : K \text{ compact } \subset E\} \tag{2.5}$$

for all Borel sets E. We say μ has limited inner regularity *if (2.5) holds for all open sets E, and for all Borel sets E with $\mu(E) < \infty$.*

Finally we say μ is regular *if μ is both outer and inner regular; and we say μ has* limited regularity *if μ is outer regular and has limited inner regulariy.*

REMARK 14. *The terminology surrounding regularity and Borel measures is* not *standardized. For example, many authors, including Rudin, say that a measure μ is a* Borel measure *if it is defined on a σ-algebra $\mathcal{A}$ that* ***contains*** *the Borel σ-algebra $\mathcal{B}$ - as opposed to identifying the measure μ with its measure space $(X, \mathcal{A}, \mu)$ and declaring it to be Borel if $\mathcal{A} = \mathcal{B}$. Rudin goes on to define μ to be regular if both (2.4) and (2.5) hold for all* Borel *sets $E \in \mathcal{B}$. Other authors insist that a regular measure satisfy the stronger requirement that (2.4) and (2.5) hold for all $E \in \mathcal{A}$. Of course, if every set $E \in \mathcal{A}$ has the form $B \cup N$ where $B \in \mathcal{B}$ is Borel and $N \in \mathcal{A}$ is null ($\mu(N) = 0$), then the two notions of regular coincide.*

We introduced the notion of *limited regularity* in order to clarify the uniqueness assertion in the Riesz representation theorem, whose statement follows.

THEOREM 32 (Riesz Representation Theorem). *Suppose that X is a locally compact Hausdorff space, and that $\Lambda : C_c(X) \to \mathbb{C}$ is a positive linear functional on $C_c(X)$. Then there is a unique positive Borel measure μ on X with limited regularity such that*

$$\Lambda f = \int_X f d\mu, \qquad f \in C_c(X). \tag{2.6}$$

Moreover, there is a σ-algebra $\mathcal{A}$ on X that contains the Borel sets in X, and an extension of μ to a measure on $\mathcal{A}$, which we continue to denote by μ, and which satisfies the following properties:

- **Local finiteness:** *$\mu(K) < \infty$ for all compact $K \subset X$,*
- **Outer $\mathcal{A}$-regularity:** *$\mu(E) = \inf\{\mu(V) : E \subset V \text{ open}\}$ for all $E \in \mathcal{A}$,*
- **Limited inner $\mathcal{A}$-regularity:** *$\mu(E) = \sup\{\mu(K) : K \text{ compact } \subset E\}$ for E open, and for $E \in \mathcal{A}$ with $\mu(E) < \infty$,*
- **Completeness:** *$A \in \mathcal{A}$ if $A \subset E \in \mathcal{A}$ and $\mu(E) = 0$.*

We will see later that inner regularity may fail for a measure μ arising in the Riesz representation theorem. On the other hand we will also see later that in nice topological spaces X, in particular those locally compact Hausdorff spaces in which every open set is a countable union of compact sets, *every* locally finite Borel measure μ is regular.

REMARK 15. *If Λ is a positive linear functional on $C_c(X)$, where X is locally compact and Hausdorff, and if μ is a positive Borel measure satisfying (2.6), then μ must be locally finite. Indeed, if K is compact, then by Urysohn's Lemma there is $f \in C_c(X)$ with $\chi_K \leq f \leq \chi_X = 1$, and so*

$$\mu(K) = \int_X \chi_K d\mu \leq \int_X f d\mu = \Lambda f < \infty.$$

Proof (of Theorem 32): We begin with the *uniqueness* of a positive Borel measure μ on X with limited regularity that satisfies the representation formula (2.6). Suppose that μ_1 and μ_2 are two such Borel measures. First we observe that because each of μ_1 and μ_2 has limited regularity, they are determined on Borel sets by their values on compact sets. Thus it suffices to prove that $\mu_1(K) = \mu_2(K)$ for all compact sets K in X.

Fix K compact and $\varepsilon > 0$. By outer regularity of μ_2 there is an open set V satisfying

$$\mu_2(V) \leq \mu_2(K) + \varepsilon.$$

By Urysohn's Lemma there is $f \in C_c(X)$ such that

$$\chi_K \leq f \leq \chi_V.$$

Altogether we thus have

$$\begin{aligned} \mu_1(K) &= \int_X \chi_K d\mu_1 \leq \int_X f d\mu_1 = \Lambda f = \int_X f d\mu_2 \\ &\leq \int_X \chi_V d\mu_2 = \mu_2(V) \leq \mu_2(K) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $\mu_1(K) \leq \mu_2(K)$, and hence also $\mu_2(K) \leq \mu_1(K)$ by symmetry.

In order to establish the *existence* of a positive Borel measure μ that satisfies the representation formula (2.6), we must work much harder. However, it turns out to be no harder to obtain the measure μ on a σ-algebra $\mathcal{A}$ with the additional properties listed in the statement of the theorem. So we now turn to proving the existence of such $\mathcal{A}$ and μ in eleven steps. Parts of the arguments below are reminiscent of some of those used in the construction of Lebesgue measure above.

We define the *support* of a complex-valued function f to be the closure of the set of x where $f(x) \neq 0$, and we denote it by $\mathsf{supp} f$; thus

$$\mathsf{supp} f \equiv \overline{\{x \in X : f(x) \neq 0\}}.$$

Step 1: For *every* subset $E \in \mathcal{P}(X)$ we define

$$\Lambda^*(E) \equiv \inf_{E \subset V \text{ open}} \left\{ \sup_{0 \leq f \leq \chi_V} \Lambda f \right\},$$

where the infimum is taken over all open sets V that contain E, and the supremum in braces is taken over all nonnegative $f \in C_{c(X)}$ such that f is subordinate to V. We first observe that for G open we have the simpler formula,

$$\Lambda^*(G) = \sup_{0 \leq f \leq \chi_G} \Lambda f,$$

and hence also

$$\Lambda^*(E) = \inf_{E \subset V \text{ open}} \Lambda^*(V), \qquad E \in \mathcal{P}(X). \tag{2.7}$$

Step 2: We claim that $\Lambda^* : \mathcal{P}(X) \to [0, \infty]$ is an *outer measure*, i.e. that Λ^* is *monotone* and *countably subadditive*.

Clearly Λ^* is monotone since if $E \subset F$, then

$$\Lambda^*(E) = \inf_{E \subset V \text{ open}} \Lambda^*(V) \leq \inf_{F \subset V \text{ open}} \Lambda^*(V) = \Lambda^*(F)$$

follows since every open set V containing F also contains E. To see that Λ^* is countably subadditive, i.e.

$$\Lambda^*\left(\bigcup_{n=1\infty} E_n\right) \leq \sum_{n=1}^{\infty} \Lambda^*(E_n), \qquad \text{for all } \{E_n\}_{n=1}^{\infty} \subset \mathcal{P}(X), \tag{2.8}$$

we first show that

$$\Lambda^*(U \cup V) \leq \Lambda^*(U) + \Lambda^*(V), \tag{2.9}$$

for all open sets U and V. Let $f \in C_c(X)$ satisfy $0 \leq f \leq \chi_{U \cup V}$. Apply the partition of unity Corollary 15 with $K = \mathsf{supp} f$ to obtain $g, h \in C_c(X)$ with

$$\begin{aligned} \chi_K &\leq g + h \leq \chi_{U \cup V}, \\ 0 &\leq g \leq \chi_U \text{ and } 0 \leq h \leq \chi_V. \end{aligned}$$

Then we have

$$\Lambda f = \Lambda[f(g+h)] = \Lambda(fg) + \Lambda(fh) \leq \Lambda^*(U) + \Lambda^*(V)$$

since $0 \leq fg \leq \chi_U$ and $0 \leq fh \leq \chi_V$. Since this holds for all $0 \leq f \leq \chi_{U \cup V}$, we can take the supremum over such f to get (2.9). Induction then yields the more general statement,

$$\Lambda^*\left(\bigcup_{n=1}^{N} V_n\right) \leq \sum_{n=1}^{N} \Lambda^*(V_n), \qquad \text{for all } V_n \text{ open.} \tag{2.10}$$

We may suppose that $\Lambda^*(E_n) < \infty$ in (2.8), and then given $\varepsilon > 0$, we can find open sets V_n such that $\Lambda^*(V_n) \leq \Lambda^*(E_n) + \frac{\varepsilon}{2^n}$ for each $n \geq 1$. Set $V \equiv \bigcup_{n=1}^{\infty} V_n$ and choose $f \in C_c(X)$ with $0 \leq f \leq \chi_V$. Since $\mathsf{supp} f$ is compact there is $N < \infty$ such that $0 \leq f \leq \chi_{\bigcup_{n=1}^{N} V_n}$. Altogether we have

$$\Lambda f \leq \Lambda^*\left(\bigcup_{n=1}^{N} V_n\right) \leq \sum_{n=1}^{N} \Lambda^*(V_n) \leq \sum_{n=1}^{N} \left(\Lambda^*(E_n) + \frac{\varepsilon}{2^n}\right) < \varepsilon + \sum_{n=1}^{N} \Lambda^*(E_n),$$

and taking the supremum over such f, we obtain

$$\Lambda^*(V) \leq \varepsilon + \sum_{n=1}^{\infty} \Lambda^*(E_n).$$

Since Λ^* is monotone and $\varepsilon > 0$ is arbitrary, we thus have

$$\Lambda^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \Lambda^*\left(\bigcup_{n=1}^{\infty} V_n\right) = \Lambda^*(V) \leq \sum_{n=1}^{\infty} \Lambda^*(E_n).$$

Step 3: We now define $\mathcal{A}$ and μ. Let

$$\mathcal{A}_{inner} \equiv \left\{ E \in \mathcal{P}(X) : \Lambda^*(E) < \infty \text{ and } \Lambda^*(E) = \sup_{\text{compact } K \subset E} \Lambda^*(K) \right\},$$

and

$$\mathcal{A} \equiv \{ E \in \mathcal{P}(X) : E \cap K \in \mathcal{A}_{inner} \text{ for every compact set } K \}.$$

Then we define $\mu : \mathcal{A} \to [0, \infty]$ by

$$\mu(E) = \Lambda^*(E), \qquad E \in \mathcal{A}.$$

We will eventually see that $\mathcal{A}_{inner}$ consists *exactly* of those sets $E \in \mathcal{A}$ such that $\mu(E) < \infty$. In the steps below we establish that $\mathcal{A}$ is a σ-algebra on X containing the Borel sets, and that μ is a positive measure on $\mathcal{A}$.

It will be convenient to use the shorthand notation $K \prec f$ (read K is subordinate to f) to mean K is compact, $f \in C_c(X)$ and $\chi_K \leq f \leq 1$; and to use $f \prec V$ (read f is subordinate to V) to mean V is open, $f \in C_c(X)$ and $0 \leq f \leq \chi_V$.

Step 4: If K is compact, then $K \in \mathcal{A}$ and

$$\mu(K) = \inf_{K \prec f} \Lambda f. \tag{2.11}$$

That $K \in \mathcal{A}$ is trivial, and to see (2.11) suppose that $K \prec f$ and $0 < \alpha < 1$. Then with $V_\alpha \equiv \{f > \alpha\}$ we have

$$\mu(K) \leq \Lambda^*(V_\alpha) = \sup_{g \prec V_\alpha} \Lambda g = \frac{1}{\alpha} \sup_{g \prec V_\alpha} \Lambda(\alpha g) \leq \frac{1}{\alpha} \Lambda(f),$$

since $\alpha g \leq f$ whenever $g \prec V_\alpha$. Letting $\alpha \to 1$ we obtain $\mu(K) \leq \Lambda(f)$.

If now $\varepsilon > 0$ there exists an open set V containing K such that $\Lambda^*(V) < \mu(K) + \varepsilon$. Urysohn's Lemma yields f so that $K \prec f \prec V$, and so altogether we have

$$\mu(K) \leq \Lambda(f) \leq \Lambda^*(V) < \mu(K) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain (2.11).

Step 5: If G is open, then

$$\Lambda^*(G) = \sup_{\text{compact } K \subset E} \mu(K). \tag{2.12}$$

In particular, if G is open and $\Lambda^*(G) < \infty$, then $G \in \mathcal{A}_{inner}$. To see (2.12), let $\alpha < \Lambda^*(G)$ so that there is $f \prec G$ with $\alpha < \Lambda f \leq \Lambda^*(G)$. Now $K = \mathsf{supp} f$ is compact and if W is an open set that contains K, then $f \prec W$ and hence

$$\Lambda f \leq \Lambda^*(W).$$

Since this holds for *all* such W we obtain

$$\Lambda f \leq \inf_{K \subset W \text{ open}} \Lambda^*(W) = \mu(K).$$

Altogether we have

$$\alpha < \Lambda f \leq \mu(K) \leq \Lambda^*(G),$$

and since α was any number less than $\Lambda^*(G)$ and K is compact, the proof of (2.12) is complete.

Step 6: Suppose that $\{E_i\}_{i=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{A}_{inner}$. Then

$$\Lambda^*\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \Lambda^*(E_i). \tag{2.13}$$

If in addition $\Lambda^*\left(\bigcup_{i=1}^{\infty} E_i\right) < \infty$, then $\bigcup_{i=1}^{\infty} E_i \in \mathcal{A}_{inner}$. We begin by proving (2.13) for a finite union of compact sets,

$$\mu\left(K \dot{\cup} L\right) = \mu(K) + \mu(L), \qquad K, L \text{ compact.} \tag{2.14}$$

Given $\varepsilon > 0$ there is by Urysohn's Lemma a function $f \in C_c(X)$ with $0 \leq f \leq 1$separating K and L in the sense that $f = 1$on K and $f = 0$ on L. From (2.11) in Step 4 we obtain g such that $K \cup L \prec g$ and

$$\Lambda g < \mu(K \cup L) + \varepsilon.$$

From (2.11) applied to $K \prec fg$ and $L \prec (1-f) g$ the linearity of Λ gives

$$\mu(K) + \mu(L) \leq \Lambda(fg) + \Lambda[(1-f) g] = \Lambda g < \mu(K \cup L) + \varepsilon.$$

Now we use that $\varepsilon > 0$ is arbitrary, together with the subadditivity of Λ^* in (2.8) of Step 2, to obtain (2.14).

Now we turn to proving (2.13) in full generality. By the countable subadditivity in (2.8) we may assume that $\Lambda^*(E) < \infty$ where $E = \bigcup_{i=1}^{\infty} E_i$. Recall that $E_i \in \mathcal{A}_{inner}$. Thus given $\varepsilon > 0$ there are compact sets $H_i \subset E_i$ satisfying

$$\Lambda^*(E_i) < \mu(H_i) + \frac{\varepsilon}{2^i}, \qquad 1 \leq i < \infty.$$

Now set $K_n \equiv \bigcup_{i=1}^{n} H_i$ and use (2.14) repeatedly to obtain

$$\sum_{i=1}^{n} \Lambda^*(E_i) < \sum_{i=1}^{n} \mu(H_i) + \varepsilon = \mu(K_n) + \varepsilon \leq \Lambda^*(E) + \varepsilon.$$

Letting first $\varepsilon \to 0$ and then $n \to \infty$ we obtain $\sum_{i=1}^{\infty} \Lambda^*(E_i) \leq \Lambda^*(E)$, which when combined with the countable subadditivity in (2.8), yields (2.13).

Step 7: Suppose that $E \in \mathcal{A}_{inner}$ and $\varepsilon > 0$. Then there is a compact set K and an open set V such that

$$K \subset E \subset V \text{ and } \Lambda^*(V \setminus K) < \varepsilon.$$

Indeed, the definition of Λ^* in (2.7) shows that there is an open set V such that $E \subset V$ and $\Lambda^*(V) < \Lambda^*(E) + \frac{\varepsilon}{2}$; while the definition of $\mathcal{A}_{inner}$ shows that there is a compact set K such that $K \subset E$ and $\mu(K) > \Lambda^*(E) - \frac{\varepsilon}{2}$. Now $V \setminus K$ is open and $\Lambda^*(V \setminus K) \leq \Lambda^*(V) < \infty$, so that (2.12) in Step 5 implies $V \setminus K \in \mathcal{A}_{inner}$. Then (2.13) applied to $V = K \dot{\cup} (V \setminus K)$ gives

$$\Lambda^*(V \setminus K) = \Lambda^*(V) - \mu(K) < \Lambda^*(E) + \frac{\varepsilon}{2} - \left(\Lambda^*(E) - \frac{\varepsilon}{2}\right) = \varepsilon.$$

Step 8: If $A, B \in \mathcal{A}_{inner}$, then

$$A \setminus B,\ A \cap B,\ A \cup B \in \mathcal{A}_{inner}.$$

Given $\varepsilon > 0$, the previous step shows that there are compact sets K and L and open sets U and V such that

$$\begin{aligned} &K \subset A \subset U \text{ and } L \subset B \subset V, \\ &\Lambda^* (U \setminus K), \ \Lambda^* (V \setminus L) < \varepsilon. \end{aligned}$$

Then monotonicity and subadditivity (2.8) give

$$\begin{aligned} \Lambda^* (A \setminus B) &\leq \Lambda^* (U \setminus L) \\ &\leq \Lambda^* (U \setminus K) + \Lambda^* (K \setminus V) + \Lambda^* (V \setminus L) \\ &< \varepsilon + \Lambda^* (K \setminus V) + \varepsilon. \end{aligned}$$

Now $J = K \setminus V$ is a compact subset of $A \setminus B$, so we conclude that

$$\Lambda^* (A \setminus B) = \sup_{\text{compact } J \subset E} \Lambda^* (J),$$

which implies that $A \setminus B \in \mathcal{A}_{inner}$ by the definition in Step 3.

Now

$$A \setminus (A \setminus B) = A \cap (A \cap B^c)^c = A \cap (A^c \cap B) = A \cap B$$

shows that $A \cap B \in \mathcal{A}_{inner}$. Finally, (2.13) applied to $A \cup B = (A \setminus B) \dot{\cup} B$ yields $A \cup B \in \mathcal{A}_{inner}$.

Remark 16. *In the special case that X is compact, we have at this point in the proof established that $\mathcal{A} = \mathcal{A}_{inner}$ is a σ-algebra on X containing the Borel sets, and that Λ^* is a measure when restricted to $\mathcal{A}$. Indeed, Step 5 shows $\mathcal{A}_{inner}$ contains all open sets, Step 8 shows that $\mathcal{A}_{inner}$ is closed under complementation, and Step 6 then shows that $\mathcal{A}_{inner}$ is closed under countable unions - after expressing a countable union as a countable union of pairwise disjoint sets in $\mathcal{A}_{inner}$. It now follows that $\mathcal{A} = \mathcal{A}_{inner}$. The countable additivity of $\mu = \Lambda^*$ on $\mathcal{A}$ follows from Step 6.*

Step 9: $\mathcal{A}$ is a σ-algebra on X containing the Borel sets. First we show that $\mathcal{A}$ is closed under complementation. If $A \in \mathcal{A}$ and K is compact, then both K and $A \cap K$ are in $\mathcal{A}_{inner}$ and so by Step 8 we have

$$A^c \cap K = K \setminus (A \cap K) \in \mathcal{A}_{inner},$$

and this shows that $A^c \in \mathcal{A}$.

Now suppose that $A = \bigcup_{i=1}^{\infty} A_i$ where each $A_i \in \mathcal{A}$, and let K be compact. We now write $A \cap K$ as a *pairwise disjoint* union by setting

$$\begin{aligned} B_1 &= A_1 \cap K, \\ B_2 &= (A_2 \cap K) \setminus B_1, \\ &\vdots \\ B_n &= (A_n \cap K) \setminus \left(\bigcup_{i=1}^{n-1} B_i \right), \\ &\vdots \end{aligned}$$

By Step 8 and induction on n we have $B_n \in \mathcal{A}_{inner}$ for all $n \geq 1$. Then Step 6 yields

$$A \cap K = \dot{\bigcup_{n \geq 1}} B_n \in \mathcal{A}_{inner}.$$

Since this holds for all compact K we have $A \in \mathcal{A}$.

Finally, if F is a closed set, then $F \cap K$ is compact, hence $F \cap K \in \mathcal{A}_{inner}$. This proves that $F \in \mathcal{A}$ and it follows that $\mathcal{A}$ contains all the Borel sets.

Step 10: $\mathcal{A}_{inner} = \mathcal{A} \cap \{E \in \mathcal{P}(X) : \Lambda^*(E) < \infty\}$ and μ is a measure on $\mathcal{A}$. If $E \in \mathcal{A}_{inner}$ then $E \cap K \in \mathcal{A}_{inner}$ for all compact K by Steps 4 and 8. This shows that

$$\mathcal{A}_{inner} \subset \mathcal{A} \cap \{E : \Lambda^*(E) < \infty\}.$$

We can now write $\mu(E) = \Lambda^*(E)$ for $E \in \mathcal{A}_{inner}$, and in particular by Step 5, for E open and $\Lambda^*(E) < \infty$.

Conversely, suppose that $E \in \mathcal{A}$ and $\mu(E) < \infty$. Given $\varepsilon > 0$ there is an open set $V \supset E$ with $\mu(V) = \Lambda^*(V) < \infty$, hence $V \in \mathcal{A}_{inner}$. Now by Steps 5 and 7 there is a compact set $K \subset V$ with $\mu(V \setminus K) < \varepsilon$. Since $E \cap K \in \mathcal{A}_{inner}$ by definition of $\mathcal{A}$, there is by definition of $\mathcal{A}_{inner}$, a compact set $H \subset E \cap K$ with

$$\mu(E \cap K) < \mu(H) + \varepsilon.$$

By subadditivity we thus have

$$\mu(E) \leq \mu(E \cap K) + \mu(V \setminus K) < \mu(H) + 2\varepsilon,$$

which implies that $E \in \mathcal{A}_{inner}$.

Finally, Step 6 shows that μ is countably additive on $\mathcal{A}$, i.e.

$$\mu\left(\dot{\bigcup_{1 \leq i < \infty}} E_i\right) = \sum_{i=1}^{\infty} \mu(E_i),$$

since if one of the sets E_i has infinite measure, there is nothing to prove, and otherwise $E_i \in \mathcal{A}_{inner}$ for all i.

Step 11: For every $f \in C_c(X)$ we have

$$\Lambda f = \int_X f d\mu.$$

Since $\Lambda(f) = -\Lambda(-f)$, it suffices to prove the inequality

$$\Lambda f \leq \int_X f d\mu, \qquad \text{for all real } f \in C_c(X). \tag{2.15}$$

So let $f \in C_c(X)$ be real with support $K = \mathsf{supp} f$, and let the interval $[a, b]$ contain the compact range of f. Given $\varepsilon > 0$ choose points $\{y_i\}_{i=0}^n \subset \mathbb{R}$ such that

$$\begin{aligned} y_0 &< a < y_1 < y_2 < ... < y_n = b, \\ \triangle y_i &\equiv y_i - y_{i-1} < \varepsilon, \qquad 1 \leq i \leq n. \end{aligned}$$

Define sets E_i by

$$E_i \equiv f^{-1}((y_{i-1}, y_i]) \cap K, \qquad 1 \leq i \leq n.$$

Now f is continuous, hence Borel measurable, and thus the sets $\{E_i\}_{i=1}^n$ are pairwise disjoint Borel sets with union K. By Step 9, the definitions of $\mathcal{A}_{inner}$ and $\mathcal{A}$, and the continuity of f, there are opens sets V_i with

$$\begin{aligned} \mu(V_i) &< \mu(E_i) + \frac{\varepsilon}{n}, \qquad 1 \le i \le n, \\ f(x) &< y_i + \varepsilon, \qquad \text{for } x \in V_i, 1 \le i \le n. \end{aligned}$$

The partition of unity Corollary 15 yields functions $h_i \prec V_i$ satisfying

$$\sum_{i=1}^n h_i(x) = 1, \qquad x \in K.$$

Thus we have $f = \sum_{i=1}^n h_i f$ and (2.11) in Step 4 shows that

$$\mu(K) \le \Lambda\left(\sum_{i=1}^n h_i\right) = \sum_{i=1}^n \Lambda h_i.$$

We also have

$$\Lambda h_i \le \mu(V_i) < \mu(E_i) + \frac{\varepsilon}{n}, \qquad 1 \le i \le n.$$

Finally, we use that

$$\begin{aligned} h_i f &\le (y_i + \varepsilon) h_i, \\ y_i - \varepsilon &< f(x) \qquad \text{for } x \in E_i, \end{aligned}$$

to obtain

$$\begin{aligned} \Lambda f &= \sum_{i=1}^n \Lambda(h_i f) \le \sum_{i=1}^n (y_i + \varepsilon) \Lambda h_i \\ &= \left[\sum_{i=1}^n (|a| + y_i + \varepsilon) \Lambda h_i\right] - \left[|a| \sum_{i=1}^n \Lambda h_i\right] \\ &\le \left[\sum_{i=1}^n (|a| + y_i + \varepsilon) \left(\mu(E_i) + \frac{\varepsilon}{n}\right)\right] - [|a| \mu(K)] \\ &= \left[\sum_{i=1}^n (y_i - \varepsilon) \mu(E_i)\right] + [2\varepsilon \mu(K)] + \left[\frac{\varepsilon}{n} \sum_{i=1}^n (|a| + y_i + \varepsilon)\right] \\ &\le \int_X f d\mu + \varepsilon [2\mu(K) + |a| + b + \varepsilon]. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we obtain (2.15), and this completes the proof of the Riesz representation theorem 32.

3. Regularity of Borel measures

Recall the fourth example in Example 4, where the set X was a well-ordered set with last element ω_1, the first uncountable ordinal. A positive measure $\lambda : \mathcal{A} \to \{0, 1\}$ was defined on X there by

$$\lambda(E) = \begin{cases} 1 & \text{if} \quad E \cup \{\omega_1\} \text{ contains an uncountable compact set} \\ 0 & \text{if} \quad E^c \cup \{\omega_1\} \text{ contains an uncountable compact set} \end{cases},$$

for $E \in \mathcal{A}$, and where $\mathcal{A}$ was the σ-algebra given by

$$\mathcal{A} \equiv \left\{ \begin{array}{c} E \in \mathcal{P}(X) : \text{either } E \cup \{\omega_1\} \text{ or } E^c \cup \{\omega_1\} \\ \text{contains an uncountable compact set} \end{array} \right\}.$$

Then $V = P_{\omega_1} = X \setminus \{\omega_1\}$ is an uncountable open set with $\lambda(V) = 1$. On the other hand, if K is a compact subset of V, then K is closed and hence $\alpha \equiv lub(K) \in K$ and it follows that $K \subset P_{\alpha+1}$. Thus $K^c \supset S_\alpha = [\alpha+1, \omega_1]$ where $[\alpha+1, \omega_1]$ is an uncountable closed, hence compact, subset of X. It follows from the definition of λ that $\lambda(K) = 0$. In particular, the measure λ does *not* have limited regularity:

$$\lambda(V) = 1 \neq 0 = \sup_{\text{compact } K \subset V} \lambda(K).$$

We thus see that the measure λ *cannot* arise as one of the measures μ in the conclusion of the Riesz representation theorem 32. However, X is a compact Hausdorff space, so $C_c(X) = C(X)$, and $\Lambda_\lambda : C(X) \to \mathbb{C}$ is a positive linear functional on $C(X)$, where

$$\Lambda_\lambda f \equiv \int_X f d\lambda.$$

By the Riesz representation theorem 32, there is a positive Borel measure μ on X with limited regularity such that $\Lambda_\lambda = \Lambda_\mu$. Thus we see that $\lambda \neq \mu$ and the question arises as to what the measure μ with limited regularity looks like. We claim that

$$\mu = \delta_{\omega_1},$$

where δ_{ω_1} is the Dirac unit mass at the point ω_1 in X (see the third example in Example 4). To see this we must show that

$$\int_X f d\lambda = f(\omega_1) = \int_X f d\delta_{\omega_1}, \qquad f \in C(X).$$

The second equality here is trivial so we turn to proving the first equality. Given $\varepsilon > 0$, let $G = f^{-1}((f(\omega_1) - \varepsilon, f(\omega_1) + \varepsilon))$ be the set of $\alpha \in X$ such that

$$|f(\alpha) - f(\omega_1)| < \varepsilon.$$

Then G is open and so contains a successor set S_β for some $\beta < \omega_1$. Since $S_\beta = [\beta+1, \omega_1]$ is compact and uncountable, we have $\lambda(G) = 1$ and $\lambda(G^c) = 0$. Thus

$$\int_X f d\lambda = \int_G f d\lambda + \int_{G^c} f d\lambda = \int_G f d\lambda$$

where

$$f(\omega_1) - \varepsilon = \int_G (f(\omega_1) - \varepsilon) d\lambda \leq \int_G f d\lambda \leq \int_G (f(\omega_1) + \varepsilon) d\lambda = f(\omega_1) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we obtain $\int_X f d\lambda = f(\omega_1)$. Note that $\delta_{\omega_1}(V) = 0$.

It turns out that the main topological obstacle to regularity in this example is the existence of an open set that is *not* a countable union of compact sets (since every compact set K in the open uncountable set $V = P_{\omega_1}$ is at most countable). Indeed, our main theorem in this section is that if every open subset of X is a countable union of compact sets, then *every* locally finite Borel measure on X is regular! In order to prove this we will first give a mild topological condition on X that forces the measures arising in the Riesz representation theorem 32 to be regular. Note that when X is compact, regularity follows immediately from limited regularity since $\mu(X) < \infty$. The mild topological condition we impose is that X be *σ-compact*.

NOTATION 1. *Let X be a topological space. We say that X is* σ-compact *if $X = \cup_{n=1}^{\infty} K_n$ is a countable union of compact sets K_n. More generally, we say that a subset E is σ-compact if E is a countable union of compact sets. We say that a set A is an F_σ-set if A is a countable union of closed sets. We say that a set B is a G_δ-set if B is a countable intersection of open sets.*

THEOREM 33. *Suppose that X is a locally compact, σ-compact Hausdorff space. If $\mathcal{A}$ and μ are as in the conclusion of Theorem 32, then we have the following properties:*

(1) *Suppose $E \in \mathcal{A}$. Given $\varepsilon > 0$ there exist sets F closed and G open such that*

$$F \subset E \subset G \text{ and } \mu(G \setminus F) < \varepsilon. \tag{3.1}$$

(2) *μ is a regular measure, i.e. (2.4) and (2.5) hold for all Borel sets E, in fact for all $E \in \mathcal{A}$.*

(3) *If $E \in \mathcal{A}$, there is an F_σ-set A and a G_δ-set B such that*

$$A \subset E \subset B \text{ and } \mu(B \setminus A) < \varepsilon.$$

In particular, every $E \in \mathcal{A}$ is the union of an F_σ-set and a null set.

Proof: Let $X = \cup_{n=1}^{\infty} K_n$ where K_n is compact for all $n \geq 1$.

(1) Suppose $E \in \mathcal{A}$ and $\varepsilon > 0$. We first claim that there is an open set $G \supset E$ with $\mu(G \setminus E) < \frac{\varepsilon}{2}$. Indeed, $\mu(K_n \cap E) \leq \mu(K_n) < \infty$ and so the *outer $\mathcal{A}$-regularity* conclusion in Theorem 32 gives us an open set $G_n \supset K_n \cap E$ with

$$\mu(G_n \setminus (K_n \cap E)) = \mu(G_n) - \mu(K_n \cap E) < \frac{\varepsilon}{2^{n+1}}.$$

Thus with $G = \cup_{n=1}^{\infty} G_n$ we have

$$\mu(G \setminus E) \leq \sum_{n=1}^{\infty} \mu(G_n \setminus (K_n \cap E)) < \frac{\varepsilon}{2}.$$

Applying the same reasoning to E^c yields an open set U with $\mu(U \setminus E^c) < \frac{\varepsilon}{2}$. Then $F = U^c$ is closed and the sets F and G satisfy (3.1) since

$$\begin{aligned} \mu(G \setminus F) &= \mu(G \setminus E) + \mu(E \setminus F) \\ &= \mu(G \setminus E) + \mu(U \setminus E^c) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

(2) To see that μ satisfies (2.5) for all $E \in \mathcal{A}$, we note that every closed set F is σ-compact; $F = \cup_{n=1}^{\infty}(K_n \cap F)$. Thus (1) implies (2.5).

(3) Finally, for each $n \geq 1$ choose $F_n \subset E \subset G_n$ where F_n is closed and G_n is open and $\mu(G_n \setminus F_n) < \frac{1}{n}$. Then $A = \bigcup_{n=1}^{\infty} F_n$ is an F_σ-set and $B = \bigcap_{n=1}^{\infty} G_n$ is a G_δ-set with $A \subset E \subset B$ and

$$\mu(B \setminus A) \leq \mu(G_n \setminus F_n) < \frac{1}{n}, \qquad n \geq 1.$$

Thus $\mu(B \setminus A) = 0$.

Now we can prove that on nice topological spaces, every reasonable Borel measure is regular.

THEOREM 34. *Let X be a locally compact Hausdorff space satisfying*

$$\text{every open set is } \sigma-\text{compact.}$$

Suppose that λ is a positive Borel measure on X that is locally finite, i.e.

$$\lambda(K) < \infty \text{ for all compact sets } K.$$

Then λ is regular.

Proof: The map $\Lambda_\lambda : C_c(X) \to \mathbb{C}$, given by $\Lambda_\lambda f = \int_X f d\lambda$ for $f \in C_c(X)$, is a positive linear functional on $C_c(X)$. Thus Theorems 32 and 33 yield a positive regular Borel measure μ *satisfying (1)* of Theorem 33 such that

$$\Lambda_\lambda f = \int_X f d\mu, \qquad f \in C_c(X).$$

It remains to show that $\lambda = \mu$ under the hypotheses of our theorem. We will use Urysohn's Lemma and the Monotone Convergence Theorem for this.

Let V be open in X. By hypothesis, V is σ-compact, so $V = \cup_{n=1}^\infty K_n$ where each K_n is compact. Urysohn's Lemma yields a function $f_n \in C_c(X)$ such that

$$\chi_{K_n} \le f_n \le \chi_V, \qquad n \ge 1.$$

Let $g_m = \max_{1 \le n \le m} f_n$. Then $g_m \in C_c(X)$ and $g_m \nearrow \chi_V$ monotonically as $m \to \infty$. Thus the Monotone Convergence Theorem can be applied twice to obtain

$$\begin{aligned}\lambda(V) &= \int_X \chi_V d\lambda = \lim_{m\to\infty} \int_X g_m d\lambda = \lim_{m\to\infty} \Lambda_\lambda g_m \\ &= \lim_{m\to\infty} \Lambda_\mu g_m = \lim_{m\to\infty} \int_X g_m d\mu = \int_X \chi_V d\mu = \mu(V).\end{aligned} \tag{3.2}$$

Now fix a Borel set E. Let $\varepsilon > 0$. Since μ satisfies (1) of Theorem 33, there is a closed set F and an open set G such that $F \subset E \subset G$ and $\mu(G \setminus F) < \varepsilon$. In particular,

$$\mu(G) = \mu(F) + \mu(G \setminus F) \le \mu(E) + \varepsilon. \tag{3.3}$$

Now $V = G \setminus F$ is open and so (3.2) gives the same sort of inequality for λ:

$$\lambda(G) = \lambda(F) + \lambda(G \setminus F) = \lambda(F) + \mu(G \setminus F) \le \lambda(E) + \varepsilon. \tag{3.4}$$

REMARK 17. *Outer regularity of μ is all that is needed to obtain $\mu(G) \le \mu(E) + \varepsilon$ in (3.3). However, we have no such regularity information regarding λ, and in order to obtain (3.4), it is necessary to know that λ coincides with μ on an open set $G \setminus F$ of small μ-measure where F and G 'sandwich' E. This is why we need assertion (1) of Theorem 33 for μ, which is* stronger *than regularity of μ.*

Using (3.2) for the open set G, it follows that both

$$\begin{aligned}\lambda(E) &\le \lambda(G) = \mu(G) \le \mu(E) + \varepsilon, \\ \mu(E) &\le \mu(G) = \lambda(G) \le \lambda(E) + \varepsilon,\end{aligned}$$

and since $\varepsilon > 0$ is arbitrary, we conclude that $\lambda(E) = \mu(E)$.

4. Lebesgue measure on Euclidean spaces

We can use the Riesz representation theorem 32 to construct Lebesgue measure on the real line $\mathbb{R}$, and more generally on Euclidean space $\mathbb{R}^n$. The idea is to define a positive linear functional Λ on $C_c(\mathbb{R}^n)$ using the *Riemann* integral $\int_{\mathbb{R}^n} f(x)\,dx$:

$$\Lambda f \equiv \int_{\mathbb{R}^n} f(x)\,dx, \qquad f \in C_c(\mathbb{R}^n).$$

It turns out that for this purpose we don't need the full theory of Riemann integration, but just enough to define the integral of a function $f \in C_c(\mathbb{R}^n)$. The following is sufficient.

4.1. Limited Riemann integration. Let $\mathcal{D}_k \equiv \left\{\left[j2^k,(j+1)\,2^k\right)\right\}_{j\in\mathbb{Z}}$ be the collection of right open left closed intervals of length 2^k and left endpoint in $2^k\mathbb{Z}$. In $\mathbb{R}^n$ we consider the corresponding cubes

$$\mathcal{D}_k^n \equiv \left\{Q_j^k\right\}_{j\in\mathbb{Z}^n} \equiv \left\{\prod_{i=1}^n \left[j_i 2^k,(j_i+1)\,2^k\right)\right\}_{j=(j_1,\dots,j_n)\in\mathbb{Z}^n},$$

obtained by forming products $Q_{(j_1,\dots,j_n)}^k = Q_{j_1}^k \times Q_{j_2}^k \times \dots \times Q_{j_n}^k$ of intervals $Q_{j_i}^k$ in $\mathcal{D}_k$. A cube $Q \in \mathcal{D}_k^n$ has volume $|Q| = \left(2^k\right)^n = 2^{kn}$ and so for $f \in C_c(\mathbb{R}^n)$, we define upper and lower sums at level k by

$$\begin{aligned} U(f;k) &\equiv \sum_{Q\in\mathcal{D}_k^n} 2^{-kn} \sup_{x\in Q} f(x), \\ L(f;k) &\equiv \sum_{Q\in\mathcal{D}_k^n} 2^{-kn} \inf_{x\in Q} f(x). \end{aligned}$$

Clearly we have for $\ell > k$,

$$U(f;k) \ge U(f;\ell) \ge L(f;\ell) \ge L(f;k). \tag{4.1}$$

Now $K = \mathsf{supp} f$ is compact, hence contained in a large cube P that is a union of unit sized cubes in $\mathcal{D}_0^n$. Moreover, f is *uniformly* continuous on K, hence on $\mathbb{R}^n$, and it follows that

$$\begin{aligned} U(f;k) - L(f;k) &= \sum_{Q\in\mathcal{D}_k^n : Q\subset P} 2^{-kn} \left\{\sup_{x\in Q} f(x) - \inf_{x\in Q} f(x)\right\} \\ &\le |P| \sup_{Q\in\mathcal{D}_k^n} \left\{\sup_{x\in Q} f(x) - \inf_{x\in Q} f(x)\right\}, \end{aligned} \tag{4.2}$$

which tends to 0 as $k \to \infty$. Thus from (4.1) and (4.2) the limits of upper and lower sums exist and coincide. We define the Riemann integral of $f \in C_c(\mathbb{R}^n)$ to be

$$\int_{\mathbb{R}^n} f(x)\,dx \equiv \lim_{k\to\infty} U(f;k) = \lim_{k\to\infty} L(f;k).$$

It follows easily that this integral has the elementary properties

$$\begin{aligned} \int_{\mathbb{R}^n} (\alpha f + \beta g)(x)\,dx &= \alpha \int_{\mathbb{R}^n} f(x)\,dx + \beta \int_{\mathbb{R}^n} g(x)\,dx, \\ \text{for } f,g &\in C_c(\mathbb{R}^n) \text{ and } \alpha,\beta\in\mathbb{C}, \\ \int_{\mathbb{R}^n} f(x)\,dx &\le \int_{\mathbb{R}^n} g(x)\,dx, \qquad \text{for } f \le g. \end{aligned}$$

Thus the map $\Lambda : C_c(\mathbb{R}^n) \to \mathbb{C}$ given by $\Lambda f = \int_{\mathbb{R}^n} f(x)\,dx$ is a positive linear functional, and Theorems 32 and 33 apply to show that there is a σ-algebra $\mathcal{L}_n$ containing the Borel sets, and a positive measure λ_n on $\mathcal{L}_n$, called *Lebesgue measure*, that satisfies

(4.3)
(1) $\Lambda f = \int_{\mathbb{R}^n} f d\lambda_n$ for all $f \in C_c(\mathbb{R}^n)$;
(2) $\lambda_n(K) < \infty$ for all compact $K \subset \mathbb{R}^n$;
(3) Given $E \in \mathcal{L}_n$ and $\varepsilon > 0$, there is K compact and G open such that
$$K \subset E \subset G \text{ and } \lambda_n(G \setminus K) < \varepsilon;$$
(4) $A \in \mathcal{L}_n$ if $A \subset E \in \mathcal{L}_n$ and $\mu(E) = 0$.

It is now an easy matter to establish the additional properties expected of Lebesgue measure λ_n on $\mathbb{R}^n$:

(4.4)
(5) $\lambda_n([a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]) = \prod_{j=1}^{n} (b_j - a_j)$,
(6) $E + x \in \mathcal{L}_n$ and $\lambda_n(E + x) = \lambda_n(E)$ if $E \in \mathcal{L}_n$ and $x \in \mathbb{R}^n$.

EXERCISE 3. *Prove both (4.3) and (4.4).*

We have already produced in Theorem 20 an example of a subset E of the interval $[0, 1)$ that is not Lebesgue measurable, i.e. $E \in \mathcal{P}(\mathbb{R}) \setminus \mathcal{L}_1$. We can lift this example to higher dimensions simply by considering the set $E_n \equiv E \times \mathbb{R}^{n-1}$ in $\mathbb{R}^n$ since it is easy to see that $E_n \in \mathcal{P}(\mathbb{R}^n) \setminus \mathcal{L}_n$.

Let $\mathcal{B}_n$ denote the Borel σ-algebra on $\mathbb{R}^n$. Then $\mathcal{B}_n \subset \mathcal{L}_n$, and the question arises as to whether or not $\mathcal{L}_n \setminus \mathcal{B}_n$ is nonempty. In fact, we have already produced in Example 3 a subset B of the unit interval $[0, 1]$ that is not measurable, and with the additional property that there is a homeomorphism $G : [0, 1] \hookrightarrow [0, 1]$ with inverse $\Phi = G^{-1}$ such that $\Phi(B)$ is contained in the Cantor set. Since Lebesgue measure is complete, $\Phi(B) \in \mathcal{L}_1$. But $\Phi(B)$ cannot be a Borel set since a homeomorphism takes Borel sets to Borel sets! Indeed, if we extend the bijection $G : [0, 1] \hookrightarrow [0, 1]$ to a bijection $G : \mathcal{P}([0, 1]) \hookrightarrow \mathcal{P}([0, 1])$ in the natural way, $E \to G(E)$, then the pushforward of a σ-algebra is again a σ-algebra by Remark 13. Since a homeomorphism takes open sets to open sets, it follows that $\mathcal{B}_1$, the smallest σ-algebra containing the open sets, is taken under the map G to the smallest σ-algebra containing the open sets, $\mathcal{B}_1$. Thus we have shown that $\mathcal{L}_1 \setminus \mathcal{B}_1 \neq \emptyset$.

However, it turns out that the set of Lebesgue measurable sets has much larger cardinality than the set of Borel measurable sets, and we now turn to establishing this.

4.2. Cardinality of Borel sets. Recall that $\mathcal{B}_n$ is the Borel σ-algebra on $\mathbb{R}^n$. Here we show that the cardinality $|\mathcal{B}_n|$ of $\mathcal{B}_n$ is at most $2^{\omega_0} = |\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$, the cardinality of both the real numbers $\mathbb{R}$ and the power set of the natural numbers $\mathbb{N}$. On the other hand, the cardinality of the Lebesgue σ-algebra $\mathcal{L}_n$ is at least the cardinality of the power set $\mathcal{P}(E)$ of the Cantor set (since $\lambda_n(E) = 0$ and λ_n is complete). But E has cardinality 2^{ω_0}, and so

$$|\mathcal{L}_n| \geq 2^{2^{\omega_0}} > 2^{\omega_0} \geq |\mathcal{B}_n|. \tag{4.5}$$

In particular, this shows that $\mathcal{L}_n \setminus \mathcal{B}_n \neq \emptyset$. In fact there are many more Lebesgue measurable sets than Borel measurable sets in the sense of cardinality.

It turns out that $|\mathcal{L}_n| = 2^{2^{\omega_0}}$ and $|\mathcal{B}_n| = 2^{\omega_0}$, but we will content ourselves with proving only the inequalities used in (4.5). The first two inequalities are easy. To show that $|\mathcal{B}_n| \leq 2^{\omega_0}$, we start with the fact that $\mathbb{R}^n$ has a *countable* base $\mathfrak{B}$ of balls, e.g. the collection of all balls with rational radii and centers having rational coordinates. Since every open set G in $\mathbb{R}^n$ is a union of balls from $\mathfrak{B}$, namely

$$G = \bigcup \{B \in \mathfrak{B} : B \subset G\},$$

we see that $\mathcal{G} \equiv \{G \in P(\mathbb{R}^n) : G \text{ is open}\}$ has cardinality $|\mathcal{G}| \leq 2^{\omega_0}$, and so also

$$|\mathcal{F}| \leq 2^{\omega_0}, \text{ where } \mathcal{F} \equiv \{F \in \mathcal{P}(\mathbb{R}^n) : F \text{ is closed}\}.$$

Now we consider the $\sigma - \delta$ operator $\Delta\Sigma$ that maps $\mathcal{P}(\mathbb{R}^n)$ to itself by

$$\Delta\Sigma\mathcal{E} \equiv \left\{ \bigcap_{\ell=1}^{\infty} \bigcup_{k=1}^{\infty} E_k^\ell : E_k^\ell \in \mathcal{E} \text{ for all } k, \ell \geq 1 \right\}.$$

We apply $\Delta\Sigma$ iteratively to the set $\mathcal{F}$ to obtain larger and larger 'sets of sets':

$$\mathcal{F}_0 \equiv \mathcal{F}, \quad \mathcal{F}_1 \equiv \Delta\Sigma\mathcal{F}, \quad \mathcal{F}_m \equiv (\Delta\Sigma)^m \mathcal{F} = \Delta\Sigma\mathcal{F}_{m-1}, \quad \text{for } m \geq 1.$$

At this point we assume minimal familiarity with *ordinal arithmetic*. Then we can continue with transfinite induction to define $\mathcal{F}_\alpha$ inductively for every ordinal $\alpha \leq \omega_1$, where ω_1 is the first uncountable ordinal:

$$\mathcal{F}_\alpha \equiv \begin{cases} \Delta\Sigma\mathcal{F}_{\alpha-1} & \text{if } \quad \alpha \text{ is a successor ordinal} \\ \bigcup_{\beta<\alpha} \mathcal{F}_\beta & \text{if } \quad \alpha \text{ is a limit ordinal} \end{cases}, \qquad \alpha \leq \omega_1.$$

One easily sees that $|\mathcal{F}_\alpha| \leq 2^{\omega_0}$ for all $\alpha < \omega_1$ by transfinite induction. Then we have

$$|\mathcal{F}_{\omega_1}| = \left| \bigcup_{\alpha<\omega_1} \mathcal{F}_\alpha \right| \leq \omega_1 \cdot 2^{\omega_0} \leq 2^{\omega_0} \cdot 2^{\omega_0} = 2^{\omega_0}. \tag{4.6}$$

CLAIM 5. $\mathcal{F}_{\omega_1} = \mathcal{B}_n$.

It follows immediately from (4.6) and the claim that $|\mathcal{B}_n| \leq 2^{\omega_0}$, and this completes our proof of (4.5).

Proof of Claim: We first use transfinite induction to show that $\mathcal{F}_{\omega_1} \subset \mathcal{B}_n$. Indeed, fix $\alpha \leq \omega_1$ and suppose that $\mathcal{F}_\beta \subset \mathcal{B}_n$ for all $\beta < \alpha$. If α is a successor ordinal, then $\mathcal{F}_{\alpha-1} \subset \mathcal{B}_n$ and

$$\mathcal{F}_\alpha = \Delta\Sigma\mathcal{F}_{\alpha-1} \subset \Delta\Sigma\mathcal{B}_n \subset \mathcal{B}_n.$$

If α is a limit ordinal, then $\mathcal{F}_\beta \subset \mathcal{B}_n$ for all $\beta < \alpha$ implies that

$$\mathcal{F}_\alpha = \bigcup_{\beta<\alpha} \mathcal{F}_\beta \subset \mathcal{B}_n.$$

Conversely, we begin by showing that the collection $\mathcal{F}_{\omega_1}$ is closed under countable unions. Suppose that $\{E_m\}_{m=1}^\infty \subset \mathcal{F}_{\omega_1}$. Then for each m, the set $E_m \in \mathcal{F}_{\alpha_m}$ for some $\alpha_m < \omega_1$. Now

$$\alpha \equiv \sup_{m\geq 1} \alpha_m < \omega_1,$$

and so $E_m \in \mathcal{F}_\alpha$ for all $m \geq 1$, hence

$$\bigcup_{m=1}^{\infty} E_m \in \Delta\Sigma\mathcal{F}_\alpha = \mathcal{F}_{\alpha+1} \subset \mathcal{F}_{\omega_1}.$$

Next we show that $\mathcal{F}_{\omega_1}$ is closed under complementation. For this we use the complementation operator $\Theta : \mathcal{P}(\mathbb{R}^n) \to \mathcal{P}(\mathbb{R}^n)$ defined by

$$\Theta\mathcal{E} \equiv \{E^c : E \in \mathcal{E}\}.$$

Suppose now that $F \in \mathcal{F}_0 = \mathcal{F}$. Then $F^c \in \Delta\Sigma\mathcal{F}$ since F^c is open and every open set is an F_σ-set. Thus we have $\Theta\mathcal{F}_0 \subset \mathcal{F}_1$. We can now prove by transfinite induction that

$$\Theta\mathcal{F}_\alpha \subset \mathcal{F}_{2\alpha+1}, \qquad \text{for all ordinals } \alpha < \omega_1. \tag{4.7}$$

Indeed, fix an ordinal $\alpha < \omega_1$ and make the induction assumption that $\Theta\mathcal{F}_\beta \subset \mathcal{F}_{2\beta+1}$ for all $\beta < \alpha$. Let $E \in \mathcal{F}_\alpha$. If α is a successor ordinal, then $E = \bigcap_{\ell=1}^{\infty}\bigcup_{k=1}^{\infty} E_k^\ell$ where $E_k^\ell \in \mathcal{F}_{\alpha-1}$ for all $k, \ell \geq 1$. The induction assumption applies to $\alpha - 1$ and we obtain

$$\begin{aligned} E^c &= \left(\bigcap_{\ell=1}^{\infty}\bigcup_{k=1}^{\infty} E_k^\ell\right)^c = \bigcup_{k=1}^{\infty}\bigcap_{\ell=1}^{\infty}\left(E_k^\ell\right)^c \in (\Delta\Sigma)^2\,\Theta\mathcal{F}_{\alpha-1} \\ &\subset (\Delta\Sigma)^2\,\mathcal{F}_{2(\alpha-1)+1} = (\Delta\Sigma)^2\,\mathcal{F}_{2\alpha-1} = \mathcal{F}_{2\alpha+1}. \end{aligned}$$

If α is a limit ordinal, then $E \in \mathcal{F}_\beta$ for some $\beta < \alpha$. The induction assumption applies to β and we obtain $E^c \in \Theta\mathcal{F}_\beta \subset \mathcal{F}_{2\beta+1} \subset \mathcal{F}_{2\alpha+1}$. This completes the proof of (4.7). Finally, if $E \in \mathcal{F}_{\omega_1}$, then $E \in \mathcal{F}_\alpha$ for some $\alpha < \omega_1$ and since $2\alpha + 1 < \omega_1$, we have from (4.7) that

$$E^c \in \mathcal{F}_{2\alpha+1} \subset \mathcal{F}_{\omega_1}.$$

Altogether, we have shown that $\mathcal{F}_{\omega_1}$ is a σ-algebra on $\mathbb{R}^n$ containing the closed sets $\mathcal{F}$. Thus $\mathcal{F}_{\omega_1} \supset \mathcal{B}_n$ since $\mathcal{B}_n$ is the *smallest* σ-algebra on $\mathbb{R}^n$ containing $\mathcal{F}$. This completes the proof of the claim.

5. Littlewood's three principles

A valuable quote from J. E. Littlewood is this:

Quote: *"The extent of knowledge required is nothing like so great as is sometimes supposed. There are three principles, roughly expressible in the following terms: Every [measurable] set is nearly a finite union of intervals; every [measurable] function is nearly continuous; every convergent sequence of [measurable] functions is nearly uniformly convergent. Most of the results of [the theory] are fairly intuitive applications of these ideas, and the student armed with them should be equal to most occasions when real variable theory is called for. If one of the principles would be the obvious means to settle the problem if it were 'quite' true, it is natural to ask if the 'nearly' is near enough, and for a problem that is actually solvable it generally is."*

In this quote, Littlewood is referring to Lebesgue measure on the real line, but the principles apply with little change to regular measures as well.

Littlewood's **first principle** is embodied in Theorems 33 and 34 for regular measures. In the case of Lebesgue measure, it is explicitly contained in property (3) of (4.3):

(3) Given $E \in \mathcal{L}_1$ and $\varepsilon > 0$, there is K compact and G open such that
$$K \subset E \subset G \text{ and } \lambda_1(G \setminus K) < \varepsilon.$$

Indeed, since G is an open subset of $\mathbb{R}$, it follows that $G = \dot{\bigcup}_{n=1}^{\infty} I_n$ is an at most countable union of pairwise disjoint intervals I_n. Choose $N < \infty$ such that

$$\lambda_1 (G \setminus K) + \sum_{n=N+1}^{\infty} \lambda_1 (I_n) < \varepsilon.$$

Then if we define the *symmetric difference* of two sets E and F by

$$\triangle (E, F) = (E \setminus F) \cup (F \setminus E),$$

we have

$$\begin{aligned} \triangle \left(\dot{\bigcup}_{n=1}^{N} I_n, E \right) &= \left(\left(\dot{\bigcup}_{n=1}^{N} I_n \right) \setminus E \right) \cup \left(E \setminus \left(\dot{\bigcup}_{n=1}^{N} I_n \right) \right) \\ &\subset (G \setminus K) \cup \left(\dot{\bigcup}_{n=N+1}^{\infty} I_n \right), \end{aligned}$$

and so

$$\lambda_1 \left(\triangle \left(E, \dot{\bigcup}_{n=1}^{N} I_n \right) \right) \leq \lambda_1 (G \setminus K) + \sum_{n=N+1}^{\infty} \lambda_1 (I_n) < \varepsilon,$$

which is what Littlewood meant by "*Every [measurable] set is nearly a finite union of intervals*".

Littlewood's **second principle** is embodied in Lusin's theorem.

THEOREM 35 (Lusin's Theorem). *Suppose that X is a locally compact Hausdorff space, and that μ is a measure on a σ-algebra $\mathcal{A}$ that satisfies the four properties in the conclusion of the Riesz representation theorem 32, namely* local finiteness*,* outer $\mathcal{A}$-regularity*,* limited inner $\mathcal{A}$-regularity*, and* completeness*. Suppose also that $f : X \to \mathbb{C}$ is measurable and that f vanishes outside a measurable set E of finite measure. Then given $\varepsilon > 0$, there is $g \in C_c(X)$ such that both*

$$\begin{aligned} &\mu (\{x \in X : f(x) \neq g(x)\}) < \varepsilon, \\ &\sup_{x \in X} |g(x)| \leq \sup_{x \in X} |f(x)|. \end{aligned} \tag{5.1}$$

The following theorem of Tietze, whose proof is deferred until after we have used it to prove Lusin's theorem, is the key to our proof of Lusin's theorem.

THEOREM 36 (Tietze extension theorem). *Suppose that X is a locally compact Hausdorff space, A is a closed subset of X, and that $f : A \to \mathbb{R}$ is continuous with compact support. Then there is a continuous extension $g : X \to \mathbb{R}$ satisfying both*

$$\begin{aligned} g(x) &= f(x), \qquad x \in A, \\ \sup_{x \in X} |g(x)| &\leq \sup_{x \in A} |f(x)|. \end{aligned}$$

We may take $g \in C_c(X)$.

Proof (of Lusin's Theorem): We first claim that it suffices to prove Lusin's theorem for *real-valued* functions. Indeed, suppose Lusin's theorem holds for real-valued functions, and let $f = u + iv$ where u and v are real-valued. We may assume that $0 < R \equiv \sup_{x \in X} |f(x)| < \infty$ since otherwise the complex-valued case follows immediately from the real-valued case. Now u and v are both measurable, and so there are real-valued functions $\varphi, \psi \in C_c(X)$ with

$$\mu(\{x \in X : u(x) \neq \varphi(x)\}) + \mu(\{x \in X : v(x) \neq \psi(x)\}) < \varepsilon.$$

Now define

$$g(x) = \begin{cases} \varphi(x) + i\psi(x) & \text{if} \quad |\varphi(x) + i\psi(x)| \leq R \\ \frac{\varphi(x)+i\psi(x)}{|\varphi(x)+i\psi(x)|} R & \text{if} \quad |\varphi(x) + i\psi(x)| \geq R \end{cases}.$$

Then $g \in C_c(X)$ and satisfies (5.1).

Now suppose that f is real-valued and measurable on X. By outer $\mathcal{A}$-regularity and limited inner $\mathcal{A}$-regularity of μ we can choose K compact and G open such that $K \subset E \subset G$ and

$$\mu(G \setminus K) = \mu(G) - \mu(K) < \frac{\varepsilon}{2}.$$

Let $\{B_n\}_{n=1}^{\infty}$ be a countable base of open intervals for $\mathbb{R} \setminus \{0\}$. Then for each $n \geq 1$, $f^{-1}(B_n)$ is a measurable subset of E and so by outer $\mathcal{A}$-regularity and limited inner $\mathcal{A}$-regularity of μ, there are open sets G_n and compact sets K_n such that

$$K_n \subset f^{-1}(B_n) \subset G_n,$$
$$\mu(G_n \setminus K_n) < \frac{\varepsilon}{2^{n+1}}.$$

Now let

$$A = G^c \cup \left\{ K \setminus \left(\bigcup_{n=1}^{\infty} (G_n \setminus K_n) \right) \right\}.$$

Then A is a closed set and the restriction $f_A : A \to \mathbb{R}$ of f to A is continuous and has compact support. Indeed, $\mathsf{supp} f_A$ is contained in K, and hence is compact. Moreover,

$$(f_A)^{-1}(B_n) = f^{-1}(B_n) \cap A = G_n \cap A$$

is relatively open in A for each $n \geq 1$, and

$$f^{-1}(0) \cap A = G^c = K^c \cap A$$

is relatively open in A as well. It follows easily that f_A is continuous. The Tietze extension theorem now yields $g : X \to \mathbb{R}$ continuous with compact support, and such that $f = g$ on A and $\sup_X |g| \leq \sup_A |f|$. Moreover,

$$\mu(A^c) \leq \mu(G \setminus K) + \mu\left(\bigcup_{n=1}^{\infty} (G_n \setminus K_n) \right) < \frac{\varepsilon}{2} + \sum_{n=1}^{\infty} \frac{\varepsilon}{2^{n+1}} = \varepsilon.$$

Proof (of the Tietze extension theorem): Let $R = \sup_{x \in A} |f(x)|$. Then $R < \infty$ since $\mathsf{supp} f$ is compact, and f is continuous. We may suppose that $R > 0$ as well and, upon replacing f with $\frac{f}{R}$, we may suppose that $R = 1$. Thus $f : A \to [-1, 1]$ is continuous. Define

$$B = \left\{ x \in A : f(x) \leq -\frac{1}{3} \right\} \text{ and } C = \left\{ x \in A : f(x) \geq \frac{1}{3} \right\}.$$

Then B and C are compact sets since $\mathsf{supp} f$ is compact by hypothesis. Urysohn's Lemma now yields $g_1 \in C_c(X)$ with $\chi_B \leq g_1 \leq \chi_{C^c}$, and so $f_1 = \frac{2}{3}\left(g_1 - \frac{1}{2}\right)$ is continuous (but no longer compactly supported) and satisfies $|f_1(x)| \leq \frac{1}{3}$ for all $x \in X$, as well as

$$f_1(x) = \begin{cases} -\frac{1}{3} & \text{if} \quad x \in B \\ \frac{1}{3} & \text{if} \quad x \in C \end{cases}.$$

It follows that we have

$$\begin{aligned} |f(x) - f_1(x)| &\leq \frac{2}{3} \text{ for all } x \in A, \\ |f_1(x)| &\leq \frac{1}{3} \text{ for all } x \in X. \end{aligned}$$

Indeed, to see the first inequality, simply consider the three cases $x \in B$, $x \in C$ and $x \in A \setminus (B \cup C)$ separately. In order to iterate this construction, it is important to be able to take $f_1 \in C_c(X)$. To achieve this, use Urysohn's Lemma to obtain a function $h \in C_c(X)$ satisfying $\chi_{\mathsf{supp} f} \preccurlyeq h \preccurlyeq \chi_X$, and then replace f_1 with hf_1.

We now repeat this construction, but applied and rescaled to the continuous function

$$(f - f_1) : A \to \left[-\frac{2}{3}, \frac{2}{3}\right],$$

that has compact support since both f and f_1 do, to obtain a continuous function $f_2 : X \to \mathbb{R}$ satisfying

$$\begin{aligned} |(f - f_1)(x) - f_2(x)| &\leq \left(\frac{2}{3}\right)^2 \text{ for all } x \in A, \\ |f_2(x)| &\leq \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) \text{ for all } x \in X. \end{aligned}$$

Once again we can assume that $f_2 \in C_c(X)$ upon multiplying it by a function $h \in C_c(X)$ satisfying $\chi_{\mathsf{supp}(f-f_1)} \preccurlyeq h \preccurlyeq \chi_X$. We continue by induction to obtain for each $n \geq 1$ a continuous function $f_n : X \to \mathbb{R}$ with compact support satisfying

$$\begin{aligned} \left| f(x) - \sum_{j=1}^{n} f_j(x) \right| &\leq \left(\frac{2}{3}\right)^n \text{ for all } x \in A, \\ |f_n(x)| &\leq \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{n-1} \text{ for all } x \in X. \end{aligned}$$

Now the infinite series $\sum_{j=1}^{\infty} f_j$ converges uniformly on X to a continuous function g on X that satisfies

$$\begin{aligned} f(x) &= g(x) \text{ for all } x \in A, \\ \sup_{x \in X} |g(x)| &\leq \sum_{n=1}^{\infty} \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{n-1} = 1 = \sup_{x \in A} |f(x)|. \end{aligned}$$

If g is not compactly supported we may multiply it by a 'Urysohn' function $h \in C_c(X)$ satisfying $\chi_{\mathsf{supp} f} \preccurlyeq h \preccurlyeq \chi_X$. This completes the proof of the Tietze extension theorem.

Littlewood's **third principle** is embodied in Egoroff's theorem.

THEOREM 37 (Egoroff's theorem). *Suppose that $(X, \mathcal{A}, \mu)$ is a finite measure space, i.e. $\mu(X) < \infty$. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence of complex-valued measurable* *functions on X that converges pointwise at every $x \in X$. For every $\varepsilon > 0$, there is a measurable set $E \in \mathcal{A}$ satisfying*

$$\begin{aligned} &\mu(X \setminus E) < \varepsilon, \\ &\{f_n\}_{n=1}^{\infty} \text{ converges uniformly on } E. \end{aligned} \tag{5.2}$$

Proof: For every $n, k \in \mathbb{N}$ define the set

$$S(n,k) \equiv \bigcap_{i,j \geq n} \left\{ x \in X : |f_i(x) - f_j(x)| < \frac{1}{k} \right\}.$$

Momentarily fix $k \geq 1$. The sequence of sets $\{S(n,k)\}_{n=1}^{\infty}$ is nondecreasing, i.e. $S(n,k) \subset S(n+1,k)$ for all $n \geq 1$. Moreover, $\bigcup_{n=1}^{\infty} S(n,k) = X$ since $\{f_n(x)\}_{n=1}^{\infty}$ is a Cauchy sequence for every $x \in X$. It follows that

$$\lim_{n \to \infty} \mu(S(n,k)) = \mu\left(\bigcup_{n=1}^{\infty} S(n,k)\right) = \mu(X), \qquad \text{for each fixed } k \geq 1.$$

We now construct a sequence $\{n_k\}_{k=1}^{\infty}$ of positive integers such that

$$E \equiv \bigcap_{k=1}^{\infty} S(n_k, k)$$

satisfies (5.2). For each $k \geq 1$ choose n_k so large that

$$\mu(X \setminus S(n_k,k)) = \mu(X) - \mu(S(n_k,k)) < \frac{\varepsilon}{2^k}.$$

Note that the first equality above uses our assumption that $\mu(X) < \infty$. Then we have

$$\begin{aligned} \mu(X \setminus E) &= \mu\left(X \setminus \left(\bigcap_{k=1}^{\infty} S(n_k,k)\right)\right) = \mu\left(\bigcup_{k=1}^{\infty} S(n_k,k)^c\right) \\ &\leq \sum_{k=1}^{\infty} \mu(S(n_k,k)^c) < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k} = \varepsilon. \end{aligned}$$

Finally, given $\eta > 0$ choose $k > \frac{1}{\eta}$. Then for all $i, j \geq n_k$ and for all $x \in E \subset S(n_k,k)$ we have

$$|f_i(x) - f_j(x)| < \frac{1}{k} < \eta,$$

which shows that $\{f_n\}_{n=1}^{\infty}$ converges *uniformly* on E.

CHAPTER 7

Lebesgue, Banach and Hilbert spaces

Let $(X, \mathcal{A}, \mu)$ be a measure space. We have already met the space of integrable complex-valued functions on X:

$$L^1(\mu) = \left\{ f : X \to \mathbb{C} : \int_X |f| \, d\mu < \infty \right\}.$$

Here the superscript 1 in $L^1(\mu)$ refers to the power of $|f|$ in the integral $\int_X |f| \, d\mu$. Using linearity and monotonicity of the integral, we see that $L^1(\mu)$ is a complex vector space:

$$\int_X |\alpha f + \beta g| \, d\mu \le \int_X (|\alpha| \, |f| + |\beta| \, |g|) \, d\mu = |\alpha| \int_X |f| \, d\mu + |\beta| \int_X |g| \, d\mu$$

is finite for all $f, g \in L^1(\mu)$ and $\alpha, \beta \in \mathbb{C}$. In fact, the integral $\int_X |f| \, d\mu$ defines a *norm* on the vector space $L^1(\mu)$ provided we identify any two functions f and g in $L^1(\mu)$ that differ only on a set of measure zero. More precisely, we declare $f \sim g$ if

$$\mu(\{x \in X : f(x) \ne g(x)\}) = 0.$$

It is easy to see that $\sim$ is an equivalence relation on $L^1(\mu)$ and that the map

$$[f] \to \int_X |f| \, d\mu$$

defines a norm on the *quotient space*

$$\mathcal{L}^1(\mu) \equiv L^1(\mu) / \sim .$$

For $f \in L^1(\mu)$, we are using the notation $[f] \in \mathcal{L}^1(\mu)$ above to denote the equivalence class containing f. Recall that a *norm* on a complex vector space V is a function

$$\|\cdot\| : V \to [0, \infty) \text{ by } x \to \|x\|$$

that satisfies

$$\begin{aligned} \|\alpha x\| &= |\alpha| \, \|x\|, \qquad \alpha \in \mathbb{C}, x \in V, \\ \|x + y\| &\le \|x\| + \|y\|, \qquad x, y \in V. \end{aligned}$$

Every norm gives rise to an associated metric d on V defined by

$$d(x, y) = \|x - y\|, \qquad x, y \in V.$$

If the metric space (V, d) is complete, we call V a *Banach space*.

In the next section, we will extend these considerations to the Lebesgue spaces $L^p(\mu)$ defined for each $0 < p \le \infty$. We will see that for $1 \le p \le \infty$, $L^p(\mu)$ is a complete normed linear space, referred to as a *Banach space*. Moreover, the special case $L^2(\mu)$ has many remarkable additional properties, and is the prototypical example of a *Hilbert space*.

But first we use the Dominated Convergence Theorem together with Lusin's Theorem to make a connection between the spaces $C_c(X)$ and $L^1(\mu)$ in the case that μ and X are related as in the Riesz representation theorem 32. In the special case that

$$\mu(V) > 0 \text{ for every open set } V, \tag{0.3}$$

$C_c(X)$ can be considered to be a subset of the space $L^1(\mu)$ of equivalence classes of integrable functions. Indeed, if $f, g \in C_c(X)$ differ only on a set of measure zero, then they differ nowhere at all. This is because $\{x \in X : f(x) \neq g(x)\}$ is an open set, and if it has μ-measure zero, then by (0.3) it is empty. On the other hand, without assuming (0.3), we can still consider the collection of equivalence classes $[f]$ of functions $f \in C_c(X)$, and we will show that this subspace is *dense* in $L^1(\mu)$.

Lemma 26. *Suppose that X is a locally compact Hausdorff space. If a σ-algebra $\mathcal{A}$ and a positive measure μ are as in the conclusion of Theorem 32, then $C_c(X)$ is* dense *in the metric space $L^1(\mu)$.*

Proof. Fix $f \in L^1(\mu)$ and $\varepsilon > 0$. For $n \in \mathbb{N}$ let

$$f_n(x) = \begin{cases} f(x) & \text{if} \quad \frac{1}{n} \leq |f(x)| \leq n \\ 0 & \text{if} \quad |f(x)| < \frac{1}{n} \text{ or } |f(x)| > n \end{cases}.$$

Then $\mu(\{x \in X : |f(x)| = \infty\}) = 0$ and so $\lim_{n\to\infty} f_n(x) = f(x)$ for μ-almost every $x \in X$. Also $|f_n(x)| \leq |f(x)|$ for all $n \in \mathbb{N}$ and $x \in X$ where $f \in L^1(\mu)$. Thus the Dominated Convergence Theorem shows that

$$\lim_{n\to\infty} \int_X |f - f_n| \, d\mu = 0,$$

and there exists n so that

$$\int_X |f - f_n| \, d\mu < \frac{\varepsilon}{2}.$$

Now f_n vanishes outside the set $\left\{|f| \geq \frac{1}{n}\right\}$, which has finite measure, and so we can use Lusin's Theorem to obtain a function $g \in C_c(X)$ such that

$$\mu(\{x \in X : f_n(x) \neq g(x)\}) < \frac{\varepsilon}{4n},$$
$$\sup_{x\in X} |g(x)| \leq \sup_{x\in X} |f_n(x)|.$$

Then

$$\sup_{x\in X} |f_n(x) - g(x)| \leq \sup_{x\in X} |f_n(x)| + \sup_{x\in X} |g(x)| \leq 2 \sup_{x\in X} |f_n(x)| \leq 2n,$$

and we have

$$\begin{aligned} dist_{L^1(\mu)}(f,g) &= \int |f - g| \, d\mu \leq \int |f - f_n| \, d\mu + \int |f_n - g| \, d\mu \\ &< \frac{\varepsilon}{2} + \int_{\{x\in X: f_n(x)\neq g(x)\}} |f_n(x) - g(x)| \, d\mu(x) \\ &\leq \frac{\varepsilon}{2} + 2n \; \mu(\{x \in X : f_n(x) \neq g(x)\}) \\ &< \frac{\varepsilon}{2} + 2n\frac{\varepsilon}{4n} = \varepsilon. \end{aligned}$$

□

1. L^p spaces

Let $(X, \mathcal{A}, \mu)$ be a measure space. For $0 < p \leq \infty$ and $f : X \to \mathbb{C}$ measurable define

$$\|f\|_{L^p(\mu)} \equiv \left(\int_X |f|^p \, d\mu \right)^{\frac{1}{p}}.$$

We denote by $L^p(\mu)$ the set of measurable functions f satisfying $\|f\|_{L^p(\mu)} < \infty$. Just as in the case $p = 1$ above, we identify functions that differ only on a set of measure zero. We will sometimes write $\|f\|_p$ instead of $\|f\|_{L^p(\mu)}$ when no confusion can arise. The next two inequalities are called *Hölder's inequality* and *Minkowski's inequality* respectively.

LEMMA 27. *Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 < p, p' < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$. Suppose that $f, g : X \to [0, \infty]$ are measurable functions. Then*

$$\begin{aligned} \int_X fg \, d\mu &\leq \|f\|_{L^p(\mu)} \|g\|_{L^{p'}(\mu)}, \\ \|f + g\|_{L^p(\mu)} &\leq \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}. \end{aligned}$$

Proof: The geometric/arithmetic mean inequality says

$$A^{\frac{1}{p}} B^{\frac{1}{p'}} \leq \frac{1}{p} A + \frac{1}{p'} B, \qquad A, B \geq 0. \tag{1.1}$$

We may assume $0 < \|f\|_{L^p(\mu)}, \|g\|_{L^{p'}(\mu)} < \infty$. Substitute $A = \frac{f(x)^p}{\|f\|^p_{L^p(\mu)}}$ and $B = \frac{g(x)^{p'}}{\|g\|^{p'}_{L^{p'}(\mu)}}$ in (1.1) and then integrate with respect to the measure μ on X to obtain

$$\begin{aligned} \int_X \frac{f(x)}{\|f\|_{L^p(\mu)}} \frac{g(x)}{\|g\|_{L^{p'}(\mu)}} d\mu(x) &\leq \int_X \left(\frac{1}{p} \frac{f(x)^p}{\|f\|^p_{L^p(\mu)}} + \frac{1}{p'} \frac{g(x)^{p'}}{\|g\|^{p'}_{L^{p'}(\mu)}} \right) d\mu(x) \\ &= \frac{1}{p} \frac{\|f\|^p_{L^p(\mu)}}{\|f\|^p_{L^p(\mu)}} + \frac{1}{p'} \frac{\|g\|^{p'}_{L^{p'}(\mu)}}{\|g\|^{p'}_{L^{p'}(\mu)}} = 1, \end{aligned}$$

which proves Hölder's inequality.

Now we apply Hölder's inequality to obtain

$$\begin{aligned} \|f + g\|^p_{L^p(\mu)} &= \int_X (f + g)^p \, d\mu \\ &= \int_X f(f + g)^{p-1} \, d\mu + \int_X g(f + g)^{p-1} \, d\mu \\ &\leq \left(\|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)} \right) \left\| (f + g)^{p-1} \right\|_{L^{p'}(\mu)}. \end{aligned} \tag{1.2}$$

However, $(p - 1) p' = p$ and so

$$\begin{aligned} \left\| (f + g)^{p-1} \right\|_{L^{p'}(\mu)} &= \left(\int_X \left[(f + g)^{p-1} \right]^{p'} d\mu \right)^{\frac{1}{p'}} \\ &= \left(\int_X (f + g)^p \, d\mu \right)^{\frac{1}{p'}} = \|f + g\|^{p-1}_{L^p(\mu)}. \end{aligned}$$

Since we may assume $0 < \|f+g\|_{L^p(\mu)} < \infty$ we can divide both sides of (1.2) by $\left\|(f+g)^{p-1}\right\|_{L^{p'}(\mu)} = \|f+g\|_{L^p(\mu)}^{p-1}$ to obtain Minkowski's inequality.

For $1 \leq p < \infty$, the subadditivity of $\|\cdot\|_{L^p(\mu)}$ shows that$L^p(\mu)$ is a *linear space*, that the function

$$f \to \|f\|_{L^p(\mu)}$$

defines a *norm* on $L^p(\mu)$, and that

$$d_{L^p(\mu)}(f,g) \equiv \|f-g\|_{L^p(\mu)}, \qquad f,g \in L^p(\mu),$$

defines a *metric* on $L^p(\mu)$. When $0 < p < 1$, $L^p(\mu)$ is still a linear space, but $\|f\|_{L^p(\mu)}$ is no longer a norm, nor is $d_{L^p(\mu)}(f,g)$ a metric. However, in this case the p^{th} power

$$\delta_{L^p(\mu)}(f,g) \equiv \|f-g\|_{L^p(\mu)}^p, \qquad f,g \in L^p(\mu),$$

defines a metric on the linear space $L^p(\mu)$ since for $A, B \geq 0$ and $0 < p < 1$,

$$(A+B)^p \leq A^p + B^p.$$

Indeed, with $B > 0$ fixed, the function $F(A) = A^p + B^p - (A+B)^p$ is increasing since $F'(A) = p\left[A^{p-1} - (A+B)^{p-1}\right] > 0$.

Exercise 4. *Under the hypotheses of Lemma 26, show that $C_c(X)$ is dense in the metric space $L^p(\mu)$.*

A key result in measure theory is the completeness of the metric space $L^p(\mu)$.

Proposition 8. *Let $(X, \mathcal{A}, \mu)$ be a measure space and $0 < p < \infty$. The metric space $L^p(\mu)$ is complete.*

Proof: We prove the case $1 \leq p < \infty$. The case $0 < p < 1$ is proved in the same way and is left to the reader. Suppose then that $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence in $L^p(\mu)$. Choose a rapidly converging subsequence $\{f_{n_k}\}_{k=1}^\infty$, by which we mean $\sum_{k=1}^\infty \left\|f_{n_{k+1}} - f_{n_k}\right\|_p < \infty$. This is easily accomplished inductively by choosing for example $\{n_k\}_{k=1}^\infty$ strictly increasing such that

$$\|f_n - f_{n_k}\|_p < \frac{1}{2^k}, \qquad n \geq n_{k+1}.$$

Then set

$$g = |f_{n_1}| + \sum_{k=1}^\infty \left|f_{n_{k+1}} - f_{n_k}\right|.$$

By the Monotone Convergence Theorem and Minkowski's inequality we have

$$\begin{aligned} \|g\|_p &= \lim_{N\to\infty} \left\{\int_X \left(|f_{n_1}| + \sum_{k=1}^N \left|f_{n_{k+1}} - f_{n_k}\right|\right)^p d\mu\right\}^{\frac{1}{p}} \\ &\leq \limsup_{N\to\infty} \left(\|f_{n_1}\|_p + \sum_{k=1}^N \left\|f_{n_{k+1}} - f_{n_k}\right\|_p\right) < \infty, \end{aligned}$$

and it follows that

$$0 \leq g(x) = |f_{n_1}(x)| + \sum_{k=1}^\infty \left|f_{n_{k+1}}(x) - f_{n_k}(x)\right| < \infty$$

for μ-almost every $x \in X$. Thus the series

$$f_{n_1}(x) + \sum_{k=1}^{\infty} \left\{ f_{n_{k+1}}(x) - f_{n_k}(x) \right\}$$

converges absolutely for μ-almost every $x \in X$ to a measurable function $f(x)$.

We claim that $f \in L^p(\mu)$ and that $\lim_{n\to\infty} f_n = f$ in $L^p(\mu)$. Indeed, Fatou's lemma gives

$$\begin{aligned}
\int_X |f(x) - f_{n_\ell}(x)|^p \, d\mu(x) &= \int_X \liminf_{k\to\infty} |f_{n_k}(x) - f_{n_\ell}(x)|^p \, d\mu(x) \\
&\leq \liminf_{k\to\infty} \int_X |f_{n_k}(x) - f_{n_\ell}(x)|^p \, d\mu(x) \\
&= \liminf_{k\to\infty} \|f_{n_k} - f_{n_\ell}\|_p^p \to 0
\end{aligned}$$

as $\ell \to \infty$ by the Cauchy condition. This shows that $f - f_{n_\ell} \in L^p(\mu)$, hence $f \in L^p(\mu)$, and also that $f_{n_\ell} \to f$ in $L^p(\mu)$ as $\ell \to \infty$. Finally, this together with the fact that $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence, easily shows that $f_n \to f$ in $L^p(\mu)$ as $n \to \infty$.

Porism 1: If $\{f_n\}_{n=1}^{\infty}$ is a rapidly converging sequence in $L^p(\mu)$,

$$\sum_{n=1}^{\infty} \|f_{n+1} - f_n\|_p < \infty,$$

then

$$\lim_{n\to\infty} f_n(x) = f_1(x) + \sum_{n=1}^{\infty} \left\{ f_{n+1}(x) - f_n(x) \right\}$$

exists for μ-almost every $x \in X$.

The completeness of $L^p(\mu)$ shows that $L^p(\mu)$ is a *Banach space* for $1 \leq p < \infty$.

2. Banach spaces

Three famous results, namely the *uniform boundedness principle*, the *open mapping theorem* and the *closed graph theorem*, hold in the generality of Banach spaces and depend on the following result of Baire.

THEOREM 38. *If X is either **(1)** a complete metric space or **(2)** a locally compact Hausdorff space, then the intersection of countably many open dense subsets of X is dense in X.*

Proof: Let $\{V_k\}_{k=1}^{\infty}$ be a sequence of open dense subsets of X, and let B_0 be any nonempty open subset of X. Define sets B_n inductively by choosing B_n open and nonempty with $\overline{B_n} \subset V_n \cap B_{n-1}$ and in addition,

$$\begin{aligned}
&diam(B_n) < \frac{1}{n} \text{ in case (1)}, \\
&\overline{B_n} \text{ is compact in case (2)}.
\end{aligned}$$

Let $K = \cap_{n=1}^{\infty} \overline{B_n}$. Then in case (1), if we choose points $x_n \in B_n$, the sequence $\{x_n\}_{n=1}^{\infty}$ is Cauchy and converges in K since each $\overline{B_n}$ is closed. Thus $K \neq \phi$. In case (2), $K \neq \phi$ since the sets $\overline{B_n}$ are compact and decreasing, hence satisfy the finite intersection property. Thus in both cases $\phi \neq K \subset B_0 \cap (\cap_{k=1}^{\infty} V_k)$, and this shows that $\cap_{k=1}^{\infty} V_k$ is dense in X.

REMARK 18. *A subset V of X is open and dense if and only if $X\setminus V$ is closed with empty interior. Thus the conclusion of Baire's Theorem can be restated as "every countable union of closed sets with empty interior in X has empty interior in X".*

DEFINITION 16. *Let E be a subset of a topological space X. We say that E is* nowhere dense *if $\overline{E}$ has empty interior, that E is of the* first category *if it is a countable union of nowhere dense sets, and that E is of the* second category *if it is not of the first category.*

Thus E is of first category *if and only if* it is a subset of a countable union of nowhere dense subsets; equivalently *if and only if* its complement E^c is a superset of a countable intersection of open dense subsets. If X is a complete metric space or a locally compact Hausdorff space, then X is of the second category. Indeed, if $X \subset \cup_{n=1}^{\infty} F_n$ where F_n are closed sets with empty interior, then

$$\phi = X^c = \left(\cup_{n=1}^{\infty} F_n\right)^c = \cap_{n=1}^{\infty} F_n^c$$

where the F_n^c are open dense sets, contradicting Baire's Theorem. Of course, the countable union of first category sets is a first category set in any topological space X, and so cannot be X if X is a complete metric space or a locally compact Hausdorff space.

2.1. The uniform boundedness principle.

THEOREM 39. *(Banach-Steinhaus uniform boundedness principle) Let X, Y be Banach spaces and Γ a set of bounded linear maps from X to Y. Let*

$$B = \left\{ x \in X : \sup_{\Lambda \in \Gamma} \|\Lambda x\|_Y < \infty \right\},$$

be the subspace of X consisting of those x with bounded Γ-orbits. If B is of the second category in X, then $B = X$ and Γ is equicontinuous, i.e.

$$\sup_{\Lambda \in \Gamma} \|\Lambda\| < \infty,$$

where $\|\Lambda\| = \sup_{\|x\| \le 1} \|\Lambda x\|_Y$.

Proof: Let $E = \cap_{\Lambda \in \Gamma} \Lambda^{-1} \left(\overline{B_Y\left(0, \frac{1}{2}\right)} \right)$ where $B_Y(0, r)$ is the ball of radius r about the origin in Y. Then $E \subset B$ and E is closed by the continuity of the maps Λ. If $x \in B$, then there is $n \in \mathbb{N}$ such that $\Lambda x \in n B_Y\left(0, \frac{1}{2}\right)$ for all $\Lambda \in \Gamma$. Thus $B = \cup_{n=1}^{\infty} nE$ and since B is of the second category in X, so is nE for some $n \in \mathbb{N}$. Since $x \to nx$ is a homeomorphism of X, we have that E is of the second category in X. Thus E has an interior point x and there is $r > 0$ so that $x - E \supset B_X(0, r)$. Then we conclude

$$\Lambda\left(B_X(0, r)\right) \subset \Lambda x - \Lambda E \subset \overline{B_Y\left(0, \frac{1}{2}\right)} - \overline{B_Y\left(0, \frac{1}{2}\right)} \subset \overline{B_Y(0, 1)},$$

which implies $\|\Lambda\| \le \frac{1}{r}$ for all $\Lambda \in \Gamma$; thus Γ is equicontinuous and $B = X$.

2.2. The open mapping theorem. A map $f : X \to Y$ where X, Y are topological spaces is *open* if $f(G)$ is open in Y for every G open in X. A famous "open mapping theorem" is that a holomorphic function f on a connected open subset Ω of the complex plane is open if it is not constant. Another is the Invariance of Domain Theorem that says $f : U \to \mathbb{R}^n$ is open if it is a continuous one-to-one map from an open set U in $\mathbb{R}^n$ into $\mathbb{R}^n$. If we consider continuous linear maps $\Lambda : X \to Y$ where X, Y are Banach spaces, then Λ is open if it is onto. Note that for a linear map $\Lambda : X \to Y$ from one normed linear space X to another Y, Λ is open if and only if $\Lambda(B_X(0,1)) \supset B_Y(0,r)$ for some $r > 0$.

THEOREM 40. *(Open mapping theorem) Suppose X, Y are Banach spaces and $\Lambda : X \to Y$ is bounded and onto. Then Λ is an open map.*

REMARK 19. *More generally, if $\Lambda : X \to Y$ is a bounded linear operator from a Banach space X to a normed linear space Y, and if ΛX is of the second category in Y, then Λ is open and onto Y, and Y is a Banach space. The proof is essentially the same as that given below.*

Proof: Since Λ is onto we have $Y = \cup_{k=1}^{\infty} \Lambda\left(kB_X\left(0, \frac{1}{4}\right)\right)$, and thus by Baire's Theorem, one of the sets $\overline{\Lambda\left(kB_X\left(0, \frac{1}{4}\right)\right)} = k\overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)}$ must have nonempty interior, and hence so must $\overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)}$, say

$$B_Y(y_0, r) \subset \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)}.$$

Then we have

$$\begin{aligned} \overline{\Lambda\left(B_X\left(0, \frac{1}{2}\right)\right)} &\supset \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right) - \Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)} \\ &\supset \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)} - \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)} \\ &\supset B_Y(y_0, r) - B_Y(y_0, r) \\ &\supset B_Y(0, r). \end{aligned} \tag{2.1}$$

It remains only to prove that $\overline{\Lambda\left(B_X\left(0, \frac{1}{2}\right)\right)} \subset \Lambda(B_X(0,1))$. For this, fix $y_1 \in \overline{\Lambda\left(B_X\left(0, \frac{1}{2}\right)\right)}$. Now the argument above shows that $\overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)}$ contains an open ball $B_Y(0, r_1)$ about the origin as well. There is $x_1 \in B_X\left(0, \frac{1}{2}\right)$ such that $\Lambda x_1 \in \Lambda\left(B_X\left(0, \frac{1}{2}\right)\right)$ satisfies $\|\Lambda x_1 - y_1\|_Y < r_1$. Then we have

$$\Lambda x_1 \in B_Y(y_1, r_1) \subset \left\{ y_1 - \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)} \right\}.$$

Now define

$$y_2 = y_1 - \Lambda x_1 \in \overline{\Lambda\left(B_X\left(0, \frac{1}{4}\right)\right)}.$$

We can repeat this procedure inductively to obtain sequences $\{x_n\}_{n=1}^{\infty} \subset X$ and $\{y_n\}_{n=1}^{\infty} \subset Y$ satisfying

$$\begin{aligned} x_n &\in B_X\left(0, \frac{1}{2^n}\right), \\ y_n &\in \overline{\Lambda\left(B_X\left(0, \frac{1}{2^n}\right)\right)}, \\ y_{n+1} &= y_n - \Lambda x_n, \end{aligned}$$

for all $n \geq 1$. Then $x = \lim_{m\to\infty} \sum_{n=1}^{m} x_n \in B_X(0,1)$ since $\|x\| \leq \sum_{n=1}^{\infty} \|x_n\| < \sum_{n=1}^{\infty} \frac{1}{2^n} = 1$, and since $\|y_n\| \leq \|\Lambda\| 2^{-n}$,

$$\Lambda x = \lim_{m\to\infty} \sum_{n=1}^{m} \Lambda x_n = \lim_{m\to\infty} \sum_{n=1}^{m} (y_n - y_{n+1}) = y_1 - \lim_{m\to\infty} y_{m+1} = y_1.$$

2.2.1. *Fourier coefficients of integrable functions.* Here we apply the Open Mapping Theorem, together with Lusin's Theorem and the Dominated Convergence Theorem, to answer a question regarding Fourier coefficients of integrable functions on the circle group $\mathbb{T}$. Recall that for $f \in L^1(\mathbb{T})$, its Fourier coefficients $\widehat{f}(n)$ are defined by

$$\widehat{f}(n) \equiv \int_0^{2\pi} f(t)\, e^{-int} \frac{dt}{2\pi}, \qquad n \in \mathbb{Z}.$$

Then

$$\left|\widehat{f}(n)\right| = \left|\int_0^{2\pi} f(t)\, e^{-int} \frac{dt}{2\pi}\right| \leq \|f\|_{L^1(\mathbb{T})}$$

for all $n \in \mathbb{Z}$, i.e. $\mathcal{F} = \widehat{\ }$ is a bounded linear map from $L^1(\mathbb{T})$ to $\ell^\infty(\mathbb{Z})$ of norm 1 ($\widehat{1} = \delta_0$). More is true because of the density of trigonometric polynomials $\sum_{n=-N}^{N} c_n e^{inx}$ in $L^1(\mathbb{T})$, namely the Riemann-Lebesgue lemma:

$$\lim_{n\to\infty} \left|\widehat{f}(n)\right| = 0, \qquad f \in L^1(\mathbb{T}).$$

Remark 20. *The set of trigonometric polynomials $\mathcal{P}$ is a* self-adjoint *subalgebra of $C(\mathbb{T})$ that* separates points *in the* compact *set $\mathbb{T}$, and is* nonvanishing *at every point of $\mathbb{T}$. Thus the Stone-Weierstrass Theorem shows that $\mathcal{P}$ is a* dense *subset of the metric space $C(\mathbb{T})$ with metric $d(f,g) \equiv \sup_{x\in\mathbb{T}} |f(x) - g(x)|$. Combining this with the density of $C(\mathbb{T})$ in $L^1(\mathbb{T})$, namely Lemma 26, we obtain that $\mathcal{P}$ is dense in $L^1(\mathbb{T})$. Indeed, given $f \in L^1(\mathbb{T})$ and $\varepsilon > 0$, choose $g \in C(\mathbb{T})$ with $\int_{\mathbb{T}} |f-g| \frac{d\theta}{2\pi} < \frac{\varepsilon}{2}$ and then choose $P \in \mathcal{P}$ such that $\sup_{\mathbb{T}} |g - P| < \frac{\varepsilon}{2}$. Altogether we have*

$$\begin{aligned} dist_{L^1(\mathbb{T})}(f,P) &= \int_{\mathbb{T}} |f-P| \frac{d\theta}{2\pi} \leq \int_{\mathbb{T}} |f-g| \frac{d\theta}{2\pi} + \int_{\mathbb{T}} |g-P| \frac{d\theta}{2\pi} \\ &< \frac{\varepsilon}{2} + \sup_{\mathbb{T}} |g-P| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

To prove the Riemann-Lebesgue lemma, simply let $\varepsilon > 0$ be given and choose $P(x) = \sum_{n=-N}^{N} c_n e^{inx}$ such $\|f - P\|_{L^1(\mathbb{T})} < \varepsilon$. Since $\widehat{P}(n) = 0$ for $|n| > N$, we have

$$\left|\widehat{f}(n)\right| = \left|\widehat{f-P}(n)\right| \leq \|f-P\|_{L^1(\mathbb{T})} < \varepsilon$$

for $|n| > N$. Thus $\mathcal{F} : L^1(\mathbb{T}) \to \ell_0^\infty(\mathbb{Z})$ with norm 1 where $\ell_0^\infty(\mathbb{Z})$ is the closed subspace of $\ell^\infty(\mathbb{Z})$ consisting of those sequences with limit zero at $\pm\infty$. The following application of the open mapping theorem shows that not every such sequence arises as the Fourier transform of an integrable function on $\mathbb{T}$.

THEOREM 41. *The Fourier transform* $\mathcal{F} : L^1(\mathbb{T}) \to \ell_0^\infty(\mathbb{Z})$ *is bounded and one-to-one, but* not *onto.*

Proof: To see that $\mathcal{F}$ is one-to-one, suppose that $f \in L^1(\mathbb{T})$ and $\widehat{f}(n) = 0$ for all $n \in \mathbb{Z}$. Then if $P(x) = \sum_{n=-N}^{N} c_n e^{inx}$ is a trigonometric polynomial,

$$\int_0^{2\pi} f(t) P(t)\, dt = \sum_{n=-N}^{N} c_n \int_0^{2\pi} f(t) e^{int} dt = 0, \tag{2.2}$$

and since trigonometric polynomials are dense in $C(\mathbb{T})$, we have

$$\int_0^{2\pi} f(t) g(t)\, dt = 0$$

for all $g \in C(\mathbb{T})$. Now let E be a measurable subset of $\mathbb{T}$. By Lusin's Theorem there is a sequence of continuous functions $\{g_n\}_{n=1}^\infty$ such that $g_n = \chi_E$ except on a set of measure at most 2^{-n} and where $\|g_n\|_\infty = 1$ for all $n \geq 1$. Thus $g_n \to \chi_E$ almost everywhere on $\mathbb{T}$, and the dominated convergence theorem shows that

$$\int_E f(t)\, dt = 0.$$

With E equal $\{t : f(t) > 0\}$ and $\{t : f(t) < 0\}$, we see that $f = 0$ a.e.

Now we prove that $\mathcal{F}$ is not onto by contradiction. If $\mathcal{R}_{\mathcal{F}} = \ell_0^\infty(\mathbb{Z})$, then the open mapping theorem shows that there is $\delta > 0$ such that

$$\left\|\widehat{f}\right\|_{\ell_0^\infty(\mathbb{Z})} \geq \delta \|f\|_{L^1(\mathbb{T})}, \qquad f \in L^1(\mathbb{T}). \tag{2.3}$$

But (2.3) fails if we take $f = \mathcal{D}_n$ for n large, since

$$\left\|\widehat{f}\right\|_{\ell_0^\infty(\mathbb{Z})} = \left\|\chi_{\{-n,1-n,\dots,n-1,n\}}\right\|_{\ell_0^\infty(\mathbb{Z})} = 1$$

while $\|\mathcal{D}_n\|_{L^1(\mathbb{T})} \nearrow \infty$.

2.3. The closed graph theorem. If X is any topological space and Y is a Hausdorff space, then every continuous map $f : X \to Y$ has a closed graph (exercise: prove this). A statement that gives conditions under which the converse holds is referred to as a "closed graph theorem". Here is an elementary example. Suppose that X and Y are metric spaces and Y is compact. If the graph of f is closed in $X \times Y$ then f is continuous. Indeed, for metric spaces it is enough to show that every sequence $\{x_n\}_{n=1}^\infty$ in X converging to a point $x \in X$ has a subsequence $\{x_{n_k}\}_{k=1}^\infty$ such that $f(x_{n_k}) \to f(x)$ as $k \to \infty$. However, since Y is compact, $\{f(x_n)\}_{n=1}^\infty$ has a convergent subsequence, say $f(x_{n_k}) \to y \in K$ as $k \to \infty$. Thus (x, y) is a limit point of the graph $G = \{(x, f(x)) : x \in X\}$, and since G is assumed closed, we have $(x, y) \in G$, i.e. $y = f(x)$. The next theorem gives the same conclusion for a *linear* map from one Banach space to another. Note that linearity is needed here since $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \begin{cases} \frac{1}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$ has a closed graph, but is not continuous at the origin.

Theorem 42. *(closed graph theorem) Suppose that X and Y are Banach spaces and $\Lambda : X \to Y$ is linear. If the graph $G = \{(x, \Lambda(x)) : x \in X\}$ is closed in $X \times Y$, then Λ is continuous.*

Proof: The product $X \times Y$ is a Banach space with the norm $\|(x,y)\| = \|x\|_X + \|y\|_Y$. Since Λ is linear and the graph G of Λ is closed, G is also a Banach space. Now the projection $\pi_1 : X \times Y \to X$ by $(x,y) \to x$ is a continuous linear map from the Banach space G *onto* the Banach space X, and the open mapping theorem thus implies that π_1 is an open map. However, π_1 is clearly one-to-one and so the inverse map $\pi_1^{-1} : X \to G$ exists and is continuous. But then the composition $\pi_2 \circ \pi_1^{-1} : X \to Y$ is also continuous where $\pi_2 : X \times Y \to Y$ by $(x,y) \to y$. We are done since $\pi_2 \circ \pi_1^{-1} = \Lambda$.

As a consequence of the closed graph theorem, we obtain the automatic continuity of symmetric linear operators on a Hilbert space.

Theorem 43. *(Hellinger and Toeplitz) Suppose that T is a linear operator on a Hilbert space H satisfying $\langle Tx, y\rangle = \langle x, Ty\rangle$ for all $x, y \in H$. Then T is continuous.*

Proof: It is enough to show that T has a closed graph G. So let (x,z) be a limit point of G. Then there is a sequence $\{x_n\}_{n=1}^\infty \subset X$ such that $x_n \to x$ and $Tx_n \to z$. For every $y \in H$ the symmetry hypothesis now shows that

$$\langle T(x_n - x), y\rangle = \langle x_n - x, Ty\rangle \to 0$$

as $n \to \infty$. But we also have

$$\langle T(x_n - x), y\rangle = \langle Tx_n, y\rangle - \langle Tx, y\rangle \to \langle z, y\rangle - \langle Tx, y\rangle$$

as $n \to \infty$. Thus $\langle z - Tx, y\rangle = 0$ for all $y \in H$ and so $z = Tx$, which shows that $(x,z) \in G$.

3. Hilbert spaces

There is a class of special Banach spaces that enjoy many of the properties of the familiar Euclidean spaces $\mathbb{R}^n$ and $\mathbb{C}^n$, namely the *Hilbert spaces*, whose norms arise from an inner product. We follow the presentation in Rudin ([**3**]).

Definition 17. *A complex vector space H is an* inner product space *if there is a map $\langle \cdot, \cdot\rangle$ from $H \times H$ to $\mathbb{C}$ satisfying for all $x, y \in H$ and $\lambda \in \mathbb{C}$,*

$$\begin{aligned} \langle x, y\rangle &= \overline{\langle y, x\rangle}, \\ \langle x + z, y\rangle &= \langle x, y\rangle + \langle z, y\rangle, \\ \langle \lambda x, y\rangle &= \lambda \langle x, y\rangle, \\ \langle x, x\rangle &\geq 0 \text{ and } \langle x, x\rangle = 0 \Longleftrightarrow x = 0. \end{aligned}$$

Then $\|x\| = \sqrt{\langle x, x\rangle}$ defines a norm on H (see below) and if this makes H into a Banach space, i.e. the metric $d(x,y) = \|x - y\|$ is complete, then we say H is a Hilbert space.

A simple example of a Hilbert space is real or complex Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$ with the usual inner product. More generally, the space $\ell^2(\mathbb{N})$ of square summable sequences $a = \{a_n\}_{n=1}^\infty$ with inner product $\langle a, b\rangle = \sum_{n=1}^\infty a_n \overline{b_n}$ is a Hilbert space. Both of these examples are included as special cases of the Hilbert space $L^2(\mu)$ where μ is a positive measure on a measure space X and the inner product is $\langle f,g\rangle = \int_X f\overline{g}d\mu$. Note that an inner product $\langle\cdot,\cdot\rangle$ on an inner product space H can always be recovered from its norm $\|\cdot\|$ by polarization:

$$4\operatorname{Re}\langle x,y\rangle = \|x+y\|^2 - \|x-y\|^2, \qquad x,y\in H.$$

LEMMA 28. *Let H be an inner product space and define $\|x\| = \sqrt{\langle x,x\rangle}$ for $x\in H$. Then $\|\cdot\|$ is a norm on H and for all $x,y\in H$,*

$$\begin{aligned} |\langle x,y\rangle| &\leq \|x\|\,\|y\|, \\ \|y\| &\leq \|\lambda x+y\| \text{ for all } \lambda\in\mathbb{C} \text{ iff } \langle x,y\rangle = 0, \\ \|x+y\|^2+\|x-y\|^2 &= 2\|x\|^2+2\|y\|^2. \end{aligned}$$

Proof: For $x,y\in H$ and $\lambda\in\mathbb{C}$,

$$0\leq \|\lambda x+y\|^2 = |\lambda|^2\|x\|^2 + 2\operatorname{Re}(\lambda\langle x,y\rangle) + \|y\|^2. \tag{3.1}$$

Thus $\langle x,y\rangle = 0$ implies $\|y\|\leq\|\lambda x+y\|$ for all $\lambda\in\mathbb{C}$. Conversely, if $x\neq 0$ we minimize the right side of (3.1) with $\lambda = -\frac{\overline{\langle x,y\rangle}}{\|x\|^2}$ to get

$$0\leq\|\lambda x+y\|^2 = -\frac{|\langle x,y\rangle|^2}{\|x\|^2}+\|y\|^2.$$

This shows that $\|y\|\leq\|\lambda x+y\|$ fails for some λ if $\langle x,y\rangle\neq 0$, and also proves the Cauchy-Schwarz inequality $|\langle x,y\rangle|\leq\|x\|\,\|y\|$. With $\lambda=1$ in (3.1) we now have

$$\begin{aligned} \|x+y\|^2 &= \|x\|^2+2\operatorname{Re}\langle x,y\rangle+\|y\|^2 \\ &\leq \|x\|^2+2\|x\|\,\|y\|+\|y\|^2 \\ &= (\|x\|+\|y\|)^2, \end{aligned}$$

which shows $\|\cdot\|$ satisfies the triangle inequality, and $\|\cdot\|$ is now easily seen to be a norm. Finally, the parallelogram law follows from expanding the inner products on the left side.

The next easy theorem lies at the heart of the great success of Hilbert spaces in analysis.

THEOREM 44. *Suppose E is a nonempty closed convex subset of a Hilbert space H. Then E contains a unique element x of minimal norm, i.e. $\|x\| = \inf_{y\in E}\|y\|$.*

Proof: Let $d=\inf_{y\in E}\|y\|$, which is finite since E is *nonempty.* Pick $\{x_n\}_{n=1}^{\infty}\subset E$ with $\|x_n\|\to d$ as $n\to\infty$. Since E is *convex*, $\frac{x_m+x_n}{2}\in E$ and so has norm at least d. The parallelogram law now yields

$$\begin{aligned} \left\|\frac{x_m-x_n}{2}\right\|^2 &= \frac{\|x_m\|^2+\|x_n\|^2}{2} - \left\|\frac{x_m+x_n}{2}\right\|^2 \\ &\leq \frac{\|x_m\|^2+\|x_n\|^2}{2} - d^2 \\ &\to \frac{d^2+d^2}{2}-d^2 = 0 \end{aligned}$$

as $m,n\to\infty$. Thus $\{x_n\}_{n=1}^{\infty}$ is Cauchy and since H is complete and E *closed*, $x=\lim_{n\to\infty}x_n\in E$. Since $\|\cdot\|$ is continuous, we have $\|x\|=d$. If $x'\in E$ also satisfies $\|x'\| = d$, then using the parallelogram law as above yields $\left\|\frac{x-x'}{2}\right\|^2 = \frac{\|x\|^2+\|x'\|^2}{2} - \left\|\frac{x+x'}{2}\right\|^2 \leq 0$, hence $x = x'$.

Let H be a Hilbert space. We say that x and y in H are perpendicular, written $x \perp y$, if $\langle x, y\rangle = 0$. We say subsets E and F of H are perpendicular, written $E \perp F$, if $\langle x, y\rangle = 0$ for all $x \in E$ and $y \in F$. Finally, we define

$$E^{\perp} = \{y \in H : \langle x, y\rangle = 0 \text{ for all } x \in E\}.$$

The next theorem uses Theorem 44 to establish an orthogonal decomposition of H relative to any closed subspace M of a Hilbert space H.

Theorem 45. *Suppose that M is a closed subspace of a Hilbert space H. Then*

$$H = M \oplus M^{\perp},$$

which means that M and $M^{\perp}$ are closed subspaces of H whose intersection is the smallest subspace $\{0\}$, and whose span is the largest subspace H. The representation

$$x = m + m^{\perp},$$

where $m \in M$ and $m^{\perp} \in M^{\perp}$, is uniquely *determined for each $x \in H$.*

Proof: $M^{\perp}$ is a subspace since $\langle x, y\rangle$ is linear in x, and is closed by the Cauchy-Schwarz inequality. The fact that $\langle x, x\rangle = 0 \Longleftrightarrow x = 0$ gives $M \cap M^{\perp} = \{0\}$. Finally, to show $M + M^{\perp} = H$, let $x \in H$ and set $E = x - M$, a nonempty closed convex set. Thus there is a unique element $m^{\perp} \in x - M$ of minimal norm having the form $x - m$ with $m \in M$. Thus for all $z \in M$ and $\lambda \in \mathbb{C}$,

$$\left\|m^{\perp}\right\| \leq \left\|m^{\perp} + \lambda z\right\|$$

and Lemma 28 implies that $\left\langle z, m^{\perp}\right\rangle = 0$ for all $z \in M$, which yields $m^{\perp} \in M^{\perp}$. Thus $x = m + m^{\perp} \in M + M^{\perp}$. If there is another such representation $x = n + n^{\perp}$, then

$$m - n = n^{\perp} - m^{\perp} \in M \cap M^{\perp} = \{0\},$$

and so $n = m$ and $n^{\perp} = m^{\perp}$.

Corollary 16. $\left(M^{\perp}\right)^{\perp} = M$.

Proof: $M \subset \left(M^{\perp}\right)^{\perp}$ is obvious, and since $M \oplus M^{\perp} = H = M^{\perp} \oplus \left(M^{\perp}\right)^{\perp}$, we cannot have that M is a proper subset of $\left(M^{\perp}\right)^{\perp}$.

Definition 18. *Let M be a closed subspace of a Hilbert space H. Define*

$$P_M : H \to M \text{ and } P_{M^{\perp}} : H \to M^{\perp}$$

by $Px = m$ and $P^{\perp}x = m^{\perp}$ where $x = m + m^{\perp}$ with $m \in M$ and $m^{\perp} \in M^{\perp}$.

Lemma 29. *P_M and $P_{M^{\perp}}$ are linear maps satisfying*

$$\|P_M x\|^2 + \|P_{M^{\perp}} x\|^2 = \|x\|^2, \qquad x \in H,$$
$$(P_M)^2 = P_M \text{ and } (P_{M^{\perp}})^2 = P_{M^{\perp}}.$$

Definition 19. *The element $P_M x$ is called the* orthogonal projection *of x onto M.*

3.0.1. *Bases.* A subset $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ of a Hilbert space H is orthonormal if $\langle u_\alpha, u_\beta \rangle = \delta_\alpha^\beta$. Given $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ orthonormal in a Hilbert space H, and $x \in H$, define the *Fourier coefficients* of x (relative to $\mathcal{U}$) by

$$\widehat{x}(\alpha) = \langle x, u_\alpha \rangle, \qquad \alpha \in A. \tag{3.2}$$

THEOREM 46. *Let $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ be an orthonormal set in a Hilbert space H, and suppose $\{a_1, ..., \alpha_N\}$ is a finite subset of A. Then*

(1) $x = \sum_{n=1}^N c_n u_{\alpha_n}$ *implies that* $c_n = \widehat{x}(\alpha_n)$ *and* $\|x\|^2 = \sum_{n=1}^N |\widehat{x}(\alpha_n)|^2$.

(2) $x \in H$ *implies*

$$\left\| x - \sum_{n=1}^N \widehat{x}(\alpha_n) u_{\alpha_n} \right\| \leq \left\| x - \sum_{n=1}^N \lambda_n u_{\alpha_n} \right\|$$

for all scalars $\lambda_1, ...\lambda_N$, and morever, equality holds if and only if $\lambda_n = \widehat{x}(\alpha_n)$ for $1 \leq n \leq N$.

(3) *The vector $\sum_{n=1}^N \widehat{x}(\alpha_n) u_{\alpha_n}$ is the orthogonal projection of x onto the linear space spanned by $\{u_{\alpha_n}\}_{n=1}^N$.*

Proof: Statement (1) is a straightforward computation using orthonormality, and (2) is equivalent, after squaring and expanding, to the inequality

$$\|x\|^2 - \sum_{n=1}^N |\widehat{x}(\alpha_n)|^2 \leq \|x\|^2 - 2\operatorname{Re} \sum_{n=1}^N \widehat{x}(\alpha_n) \overline{\lambda_n} + \sum_{n=1}^N |\lambda_n|^2,$$

which in turn follows from $\left| \sum_{n=1}^N \widehat{x}(\alpha_n) \overline{\lambda_n} \right| \leq \sqrt{\sum_{n=1}^N |\widehat{x}(\alpha_n)|^2} \sqrt{\sum_{n=1}^N |\lambda_n|^2}$. Finally, (3) follows from (2) and the definition of orthogonal projection.

THEOREM 47. *(Bessel's inequality) If $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ is an orthonormal set in a Hilbert space H, then $\sum_{\alpha \in A} |\widehat{x}(\alpha)|^2 \leq \|x\|^2$ for all $x \in H$.*

Proof: Let F be a finite subset of A and let M be the subspace spanned by $\{u_\alpha\}_{\alpha \in F}$. It is an easy exercise to use (1) of Theorem 46 to see that M is closed, and then (3) of Theorem 46 shows that $P_M x = \sum_{n=1}^N \widehat{x}(\alpha_n) u_{\alpha_n}$. Then by (1) of Theorem 46 and Lemma 29, we have

$$\sum_{\alpha \in F} |\widehat{x}(\alpha)|^2 = \left\| \sum_{n=1}^N \widehat{x}(\alpha_n) u_{\alpha_n} \right\|^2 = \|P_M x\|^2 \leq \|P_M x\|^2 + \|P_{M^\perp} x\|^2 = \|x\|^2.$$

Now take the supremum over all finite subsets F of A.

THEOREM 48. *(Riesz-Fischer) If $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ is an orthonormal set in a Hilbert space H and $\varphi \in \ell^2(A)$, then there is $x \in H$ such that $\widehat{x} = \varphi$.*

Proof: There is $E = \{\alpha_n\}_{n=1}^\infty \subset A$ such that $\varphi(\alpha) = 0$ for $\alpha \in A \diagdown E$. Then $x_N = \sum_{n=1}^N \varphi(\alpha_n) u_{\alpha_n}$ is Cauchy in H, hence convergent to some $x \in H$, and continuity now yields $\widehat{x} = \varphi$.

The following fundamental theorem regarding orthonormal sets is an easy consequence of the above results.

THEOREM 49. *Suppose $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ is an orthonormal set in a Hilbert space H. Then the following statements are equivalent:*

(1) *equality holds in Bessel's inequality, i.e.*

$$\|x\| = \left\{ \sum_{\alpha \in A} |\widehat{x}(\alpha)|^2 \right\}^{\frac{1}{2}} = \|\widehat{x}\|_{\ell^2(A)}, \qquad x \in H,$$

(2) *the linear map* $\wedge : H \to \ell^2(A)$ *defined in (3.2) is a Hilbert space isomorphism of* H *onto* $\ell^2(A)$,
(3) $\mathcal{U}$ *is a* maximal *orthonormal set (called an orthonormal* basis*)*
(4) *The linear span*

$$Span\ \mathcal{U} \equiv \left\{ \sum_{\alpha \in F} c_\alpha u_\alpha : c_\alpha \text{ scalar, } F \text{ a finite subset of } A \right\}$$

is dense in H.

Proof. We prove (1) $\Longrightarrow$ (2) $\Longrightarrow$ (3) $\Longrightarrow$ (4) $\Longrightarrow$ (1). If (1) holds, then (2) follows by the Riesz-Fischer theorem, which shows $\wedge$ is onto, and polarization, which shows that $\wedge$ preserves inner products:

$$\begin{aligned} \langle x, y \rangle_H &= \frac{1}{4} \left\{ \|x+y\|_H^2 - \|x-y\|_H^2 + i\|x+iy\|_H^2 - i\|x-iy\|_H^2 \right\} \\ &= \frac{1}{4} \left\{ \|\widehat{x}+\widehat{y}\|_{\ell^2(A)}^2 - \|\widehat{x}-\widehat{y}\|_{\ell^2(A)}^2 + i\|\widehat{x}+i\widehat{y}\|_{\ell^2(A)}^2 - i\|\widehat{x}-i\widehat{y}\|_{\ell^2(A)}^2 \right\} \\ &= \langle \widehat{x}, \widehat{y} \rangle_{\ell^2(A)}, \qquad x, y \in H. \end{aligned}$$

Now assume (2) holds Then (3) holds since otherwise, there is $v \in H$ with $\|v\|_H = 1$ such that

$$\widehat{v}(\alpha) = \langle \widehat{u_\alpha}, \widehat{v} \rangle_{\ell^2(A)} = \langle u_\alpha, v \rangle_H = 0, \qquad \alpha \in A,$$

i.e. $\widehat{v} = 0$, contradicting $\|v\|_H = \|\widehat{v}\|_{\ell^2(A)}$.

Next, assume that (3) holds. Then $Span\ \mathcal{U}$ is dense in H since otherwise, $\overline{Span\ \mathcal{U}}^\perp \neq \{0\}$ by Theorem 45, and so there is $z \in H$ with $\|z\|_H = 1$ such that $\langle z, x \rangle_H = 0$ for all $x \in Span\ \mathcal{U}$. In particular, $\langle z, u_\alpha \rangle_H = 0$ for all $\alpha \in A$, contradicting maximality of $\mathcal{U}$.

Finally, assume (4) holds so that $\overline{Span\ \mathcal{U}} = H$. The linear isometry

$$\wedge : Span\ \mathcal{U} \to \ell^2(A), \qquad \widehat{x}(\alpha) = \langle x, u_\alpha \rangle, \alpha \in A,$$

has a unique continuous extension (isometries are trivially continuous) to $\overline{Span \mathcal{U}} = H$, and this continuation is easily seen to be a linear isometry. But this is precisely (1). □

Corollary 17. *If* $\mathcal{U} = \{u_\alpha\}_{\alpha \in A}$ *is an orthonormal basis for a Hilbert space* H*, then for each* $x \in H$*, the set* $\{\alpha \in A : \widehat{x}(\alpha) \neq 0\}$ *is at most countable, i.e.*

$$\{\alpha \in A : \widehat{x}(\alpha) \neq 0\} = \{\alpha_n\}_{n=1}^{\infty \text{ or finite}},$$

and

$$x = \sum_{n=1}^{\infty \text{ or finite}} \widehat{x}(\alpha_n) u_{\alpha_n},$$

with convergence of the series in H.

PROOF. Theorem 49 (1) implies that $\sum_{\alpha\in A}|\widehat{x}(\alpha)|^2=\|x\|^2<\infty$, and it follows that

$$\{\alpha\in A:\widehat{x}(\alpha)\neq 0\}=\{\alpha_n\}_{n=1}^{\infty \text{ or finite}}$$

is at most countable. Theorem 49 (4) shows that given $\varepsilon>0$, there is an element $\sum_{n=1}^{M}\lambda_n u_{\alpha_n}$ in *Span* $\mathcal{U}$ such that $\left\|x-\sum_{n=1}^{M}\lambda_n u_{\alpha_n}\right\|<\varepsilon$, and then with $\lambda_n=0$ for $M<n\leq N$, Theorem 46 (2) shows that

$$\left\|x-\sum_{n=1}^{N}\widehat{x}(\alpha_n)u_{\alpha_n}\right\|\leq\left\|x-\sum_{n=1}^{M}\lambda_n u_{\alpha_n}\right\|<\varepsilon$$

for all $N\geq M$. □

The axiom of choice shows that there are lots of orthonormal bases in a Hilbert space.

THEOREM 50. *Every orthonormal set $\mathcal{U}$ in a Hilbert space H is contained in a maximal orthonormal set $\mathcal{V}$.*

Proof: Following the standard transfinite recipe, we let Γ be the class of all orthonormal sets containing $\mathcal{U}$, partially ordered by inclusion. By the Hausdorff Maximality Theorem, Γ contains a maximal totally ordered class Ω. It is straightforward to show that $\mathcal{V}=\cup\{\mathcal{W}:\mathcal{W}\in\Omega\}$ is a maximal orthonormal set in H.

EXAMPLE 5. *Here are two examples of orthonormal bases in the Hilbert spaces $L^2(\mathbb{T})$ and $L^2(\mathbb{R},\mu)$ respectively.*

(1) *The set $\mathcal{U}=\left\{e^{int}\right\}_{n\in\mathbb{Z}}$ is an orthonormal set in $L^2(\mathbb{T})$, i.e.*

$$\left\langle e^{imt},e^{int}\right\rangle=\int_0^{2\pi}e^{imt}\overline{e^{int}}\frac{dt}{2\pi}=\begin{cases}0 & \text{if } m\neq n\\ 1 & \text{if } m=n\end{cases}.$$

The Stone-Weierstrass Theorem, together with Exercise 4, shows that Span$\mathcal{U}$ is dense in $H=L^2(\mathbb{T})$, and thus by Theorem 49 the map $\mathcal{F}:L^2(\mathbb{T})\to\ell^2(\mathbb{Z})$ given by

$$\mathcal{F}f(n)=\widehat{f}(n)=\left\langle f,e^{int}\right\rangle=\int_0^{2\pi}f(t)\,e^{-int}\frac{dt}{2\pi},\qquad n\in\mathbb{Z},$$

is a Hilbert space isomorphism of $L^2(\mathbb{T})$ onto $\ell^2(\mathbb{Z})$. Thus $\left\{e^{int}\right\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(\mathbb{T})$.

(2) *Let $\mathcal{D}=\bigcup_{k\in\mathbb{Z}}\mathcal{D}_k$ be the union of the collections $\mathcal{D}_k\equiv\left\{\left[j2^k,(j+1)\,2^k\right)\right\}_{j\in\mathbb{Z}}$ of right open left closed intervals of length 2^k having left endpoint in $2^k\mathbb{Z}$. We refer to the intervals in $\mathcal{D}$ as* dyadic *intervals. For each dyadic interval I, the left half I_- and the right half I_+ are referred to as the* children *of I. Now suppose that μ is a positive measure on $\mathbb{R}$, and for convenience we suppose that*

$$\mu(I)>0 \text{ for every } I\in\mathcal{D}.$$

Then for every $I\in\mathcal{D}$ we define the Haar function *h_I^μ by*

$$h_I^\mu(x)=\sqrt{\frac{\mu(I_-)\mu(I_+)}{\mu(I)}}\left(-\frac{\mathbf{1}_{I_-}(x)}{\mu(I_-)}+\frac{\mathbf{1}_{I_+}(x)}{\mu(I_+)}\right),\qquad x\in\mathbb{R}.$$

Here we are writing $\mathbf{1}_{I_\pm}(x)$ *for the indicator function* $\chi_{I_\pm}(x)$*. Thus the Haar function* h_I^μ *is supported in* I *and is constant on each child* $I_\pm$ *of* I*. In the special case* μ *is Lebesgue measure* λ_1*, and* I *has length* 2^k*,* $h_I^{\lambda_1}$ *takes on the value* $-\frac{1}{\sqrt{2^{k+1}}}$ *on the left half of* I*, and the value* $\frac{1}{\sqrt{2^{k+1}}}$ *on the right half of* I *(draw a picture!). The collection of Haar functions* $\{h_I^\mu\}_{I\in\mathcal{D}}$ *has the following elementary properties:*

$$\begin{aligned}
\mathsf{supp} h_I^\mu &\subset I, \\
\int h_I^\mu d\mu &= \sqrt{\frac{\mu(I_-)\mu(I_+)}{\mu(I)}}\left\{-\frac{1}{\mu(I_-)}\int_{\mu(I_-)} d\mu + \frac{1}{\mu(I_+)}\int_{\mu(I_+)} d\mu\right\} = 0, \\
\int |h_I^\mu|^2\, d\mu &= \frac{\mu(I_-)\mu(I_+)}{\mu(I)}\left\{\frac{1}{\mu(I_-)^2}\int_{\mu(I_-)} d\mu + \frac{1}{\mu(I_+)^2}\int_{\mu(I_+)} d\mu\right\} \\
&= \frac{\mu(I_-)\mu(I_+)}{\mu(I)}\left\{\frac{1}{\mu(I_-)} + \frac{1}{\mu(I_+)}\right\} = \frac{\mu(I_+)+\mu(I_-)}{\mu(I)} = 1.
\end{aligned}$$

Moreover, there follows the crucial orthogonality property:

$$\int h_I^\mu h_J^\mu d\mu = 0, \qquad \text{if } I, J \in \mathcal{D} \text{ and } I \neq J.$$

Indeed, this follows simply from (1): $\int h_J^\mu d\mu = 0$*, and (2): if* J *is a* proper *dyadic subinterval of a dyadic interval* I*, then* h_I^μ *is constant on the support of* h_J^μ*.*

Altogether we have shown that $\{h_I^\mu\}_{I\in\mathcal{D}}$ *is an orthonormal set in* $L^2(\mu)$*. It can be shown, using the differentiation theory two chapters below, that* $\{h_I^\mu\}_{I\in\mathcal{D}}$ *is actually an orthonormal basis for* $L^2(\mu)$*.*

REMARK 21. *In the special case that* μ *is Lebesgue measure on the real line* $\mathbb{R}$*, the set of Haar functions* $\{h_I\}_{I\in\mathcal{D}}$ *is generated by translation and dilation of the* single *function*

$$h_{[0,1)} = -\mathbf{1}_{[0,\frac{1}{2})} + \mathbf{1}_{[\frac{1}{2},1)}.$$

Thus the Haar basis $\{h_I\}_{I\in\mathcal{D}}$ *is the simplest example of a* wavelet *basis, an orthonormal basis that is generated by translations and dilations of a fixed 'mother wavelet'. Such wavelet bases have been characterized, and their properties catalogued, by Daubechies and others.*

Next we give an application of Hilbert space theory and the uniform boundedness principle to *nonconvergence* of Fourier series.

3.0.2. *Nonconvergence of Fourier series of continuous functions.* Recall the orthonormal basis $\{e^{int}\}_{n\in\mathbb{Z}}$ of $L^2(\mathbb{T})$ in the example above. Now consider the symmetric partial sums $S_n f$ of the Fourier series of $f \in L^1(\mathbb{T})$:

$$\begin{aligned}
S_n f(x) &= \sum_{k=-n}^{n} \widehat{f}(k)\, e^{ikx} = \sum_{k=-n}^{n} \int_0^{2\pi} f(t)\, e^{-ikt}\frac{dt}{2\pi} e^{ikx} \\
&= \int_0^{2\pi} f(t)\left\{\sum_{k=-n}^{n} e^{ik(x-t)}\right\}\frac{dt}{2\pi} \\
&= \int_0^{2\pi} f(t)\,\mathcal{D}_n(x-t)\frac{dt}{2\pi} = f * \mathcal{D}_n(x),
\end{aligned}$$

where

$$\begin{aligned}\mathcal{D}_n(\theta) &= \sum_{k=-n}^{n} e^{ik\theta} = \frac{\left(e^{i\frac{\theta}{2}} - e^{-i\frac{\theta}{2}}\right)\sum_{k=-n}^{n} e^{ik\theta}}{e^{i\frac{\theta}{2}} - e^{-i\frac{\theta}{2}}} \\ &= \frac{e^{i\left(n+\frac{1}{2}\right)\theta} - e^{-i\left(n+\frac{1}{2}\right)\theta}}{e^{i\frac{\theta}{2}} - e^{-i\frac{\theta}{2}}} = \frac{\sin\left(n+\frac{1}{2}\right)\theta}{\sin\frac{\theta}{2}}\end{aligned}$$

satisfies

$$\begin{aligned}\int_0^{2\pi} \left|\mathcal{D}_n(\theta)\right| \frac{d\theta}{2\pi} &> 2\int_0^{\pi} \frac{\left|\sin\left(n+\frac{1}{2}\right)\theta\right|}{\left|\frac{\theta}{2}\right|} \frac{d\theta}{2\pi} \\ &= \frac{2}{\pi}\int_0^{\left(n+\frac{1}{2}\right)\pi} \left|\sin\theta\right| \frac{d\theta}{\theta} \\ &> \frac{2}{\pi}\sum_{k=1}^{n} \frac{1}{k\pi}\int_{(k-1)\pi}^{k\pi} \left|\sin\theta\right| d\theta \\ &= \frac{4}{\pi^2}\sum_{k=1}^{n}\frac{1}{k},\end{aligned}$$

and so tends to ∞ as $n \to \infty$.

From the Hilbert space theory above, we obtain that $S_n f$ converges to f in $L^2(\mathbb{T})$ for all $f \in L^2(\mathbb{T})$:

$$\left\|S_n f - f\right\|^2 = \sum_{|k|>n} \left|\widehat{f}(k)\right|^2 \to 0 \text{ as } n \to \infty, \qquad f \in L^2(\mathbb{T}).$$

For $f \in C(\mathbb{T})$ we ask if we have pointwise convergence of $S_n f$ to f on $\mathbb{T}$. However, the property $\sup_{n\geq 1} \left\|\mathcal{D}_n\right\|_{L^1(\mathbb{T})} = \infty$ of the Dirichlet kernel $\mathcal{D}_n$, when combined with the uniform boundedness principle, implies that there are continuous functions $f \in C(\mathbb{T})$ whose Fourier series $\sum_{k=-\infty}^{\infty} \widehat{f}(k) e^{ikx}$ fail to converge at some points x in $\mathbb{T}$. In fact there is a dense G_δ subset E of $C(\mathbb{T})$ (a set is a G_δ subset of X if it is a countable intersection of open subsets of X) such that $\{x \in \mathbb{T} : S_n f(x) \text{ fails to converge at } x\}$ contains a dense G_δ subset of $\mathbb{T}$ for every $f \in E$.

To see this, set $\Lambda_n f = S_n f(0) = \int_0^{2\pi} f(t)\,\mathcal{D}_n(t)\,\frac{dt}{2\pi}$. Then $\Lambda_n \in C(\mathbb{T})^*$ and $\left\|\Lambda_n\right\|^* = \int_0^{2\pi} \left|\mathcal{D}_n(t)\right| \frac{dt}{2\pi} \nearrow \infty$ as $n \to \infty$. By the uniform boundedness principle we cannot have

$$\sup_{n\geq 1} \left|\Lambda_n f\right| = \sup_{n\geq 1} \left|S_n f(0)\right| < \infty \tag{3.3}$$

for f in a dense G_δ subset of $C(\mathbb{T})$. In particular, there exists a continuous function f on $\mathbb{T}$ whose Fourier series fails to converge at 0. However, since B is a subspace, we cannot in fact have (3.3) in any open set, and it follows that

$$E_0 = \left\{ f \in C(\mathbb{T}) : \sup_{n\geq 1} \left|\Lambda_n f\right| = \infty \right\}$$

is dense. Since the map $\sup_{n\geq 1} \left|\Lambda_n f\right|$ is a lower semicontinuous function of f, we also have that E_0 is a G_δ subset.

Now choose $\{x_i\}_{i=1}^{\infty}$ dense in $\mathbb{T} = [0, 2\pi)$, and by applying the above argument with x_i in place of 0, choose E_i to be a dense G_δ subset of $C(\mathbb{T})$ such that

$$\sup_{n\geq 1} |S_n f(x_i)| = \infty, \qquad f \in E_i, \ i \geq 1.$$

By Baire's Theorem, $E = \cap_{i=1}^{n} E_i$ is also a dense G_δ subset of $C(\mathbb{T})$. Thus for every $f \in E$ we have $\sup_{n\geq 1} |S_n f(x_i)| = \infty$ for all $i \geq 1$. Now we note that $\sup_{n\geq 1} |S_n f(x)|$ is a lower semicontinuous function of x (since it is a supremum of continuous functions), and thus the set

$$\left\{ x \in \mathbb{T} : \sup_{n\geq 1} |S_n f(x)| = \infty \right\}$$

is a G_δ subset of $\mathbb{T}$ for every $f \in C(\mathbb{T})$. Combining these observations yields that there is a dense G_δ subset E of $C(\mathbb{T})$ such that for every $f \in E$, the set of x where the Fourier series of f fails to converge contains a dense G_δ subset of $\mathbb{T}$.

Remark 22. *In a complete metric space X without isolated points, every dense G_δ subset is uncountable. Indeed, if $E = \{x_k\}_{k=1}^{\infty} = \cap_{n=1}^{\infty} V_n$, V_n open, is a* count-able *dense G_δ subset of X, then $W_n = V_n \setminus \{x_k\}_{k=1}^{n}$ is still a dense open subset of X, but $\cap_{n=1}^{\infty} W_n = \phi$, contradicting Baire's Theorem.*

Remark 23. *A famous theorem of L. Carleson shows that for* every $f \in L^2(\mathbb{T})$, $\lim_{n\to\infty} S_n f(x) = f(x)$ *for a.e.* $x \in \mathbb{T}$.

4. Duality

Given any normed linear space X we define X^* to be the vector space of all continuous linear *functionals* on X, i.e. continuous linear maps $\Lambda : X \to \mathbb{C}$ (or into $\mathbb{R}$ if the scalar field is real). We recall that a map L from one normed linear space X to another Y is *linear* if $L(\lambda x + y) = \lambda Lx + Ly$ for all $x, y \in X$ and $\lambda \in \mathbb{C}$. Recall also that L is said to be *bounded* if there is a nonnegative constant C such that $\|Lx\|_Y \leq C \|x\|_X$ for all $x \in X$. The proof of the next result is easy and is left to the reader.

Lemma 30. *Let $L : X \to Y$ be linear where X, Y are normed linear spaces. Then L is bounded $\Longleftrightarrow$ L is continuous on X $\Longleftrightarrow$ L is continuous at* 0.

By Lemma 30 a linear functional is continuous on X if and only if it is continuous at the origin, or equivalently bounded. If we set

$$\|\Lambda\|^* = \sup_{\|x\|\leq 1} |\Lambda x|, \tag{4.1}$$

then it is easily verified that $\|\cdot\|^*$ is a norm on X^*, and since the scalar field is complete, so is the metric on X^* induced from $\|\cdot\|^*$. Thus X^* is a Banach space (even if X is not).

Remark 24. *Note that $\|\Lambda\|^*$ is the smallest nonnegative constant C which exhibits the boundedness of Λ on X in the inequality $|\Lambda x| \leq C \|x\|$.*

Now we specialize this definition to a Hilbert space H. An example of a continuous linear functional on H is the linear functional Λ_y associated with $y \in H$ given by

$$\Lambda_y x = \langle x, y \rangle, \qquad x \in H. \tag{4.2}$$

The boundedness of Λ_y follows from the Cauchy-Schwarz inequality $|\Lambda_y x| \leq \|y\| \, \|x\|$. In fact, this together with the choice $x = \frac{y}{\|y\|}$ in (4.1) yields $\|\Lambda_y\|^* = \|y\|$. It turns out that there are no other continuous linear functionals on H and this is the first major consequence of Theorem 45, and hence also of Theorem 44.

THEOREM 51 (Riesz representation theorem for Hilbert spaces). *Let H be a Hilbert space. Every $\Lambda \in H^*$ is of the form Λ_y for some $y \in H$. Moreover, there is a conjugate linear isometry from H to H^* given by $y \to \Lambda_y$ where Λ_y is as in (4.2).*

Proof: We've already shown that $\Lambda_y \in H^*$ with $\|\Lambda_y\|^* = \|y\|$, and since $\Lambda_{\lambda y} = \overline{\lambda}\Lambda_y$ we have that the map $y \to \Lambda_y$ is a conjugate linear isometry from H into H^*. To see that this map is onto, take $\Lambda \neq 0$ in H^* and let $\mathcal{N} = \{x \in H : \Lambda x = 0\} = \Lambda^{-1}\{0\}$ be the null space of Λ. Since $\mathcal{N}$ is a proper closed subspace of H, Theorem 45 shows that $\mathcal{N}^\perp \neq \{0\}$. Take $z \neq 0$ in $\mathcal{N}^\perp$ and note that

$$(\Lambda x)\, z - (\Lambda z)\, x \in \mathcal{N} \text{ for all } x \in H.$$

Thus

$$0 = \langle (\Lambda x)\, z - (\Lambda z)\, x, z\rangle = (\Lambda x)\, \|z\|^2 - (\Lambda z)\, \langle x, z\rangle$$

yields

$$\Lambda x = \frac{(\Lambda z)\, \langle x, z\rangle}{\|z\|^2} = \left\langle x, \frac{\overline{\Lambda z}}{\|z\|^2} z\right\rangle = \Lambda_y x, \qquad x \in H,$$

with $y = \frac{\overline{\Lambda z}}{\|z\|^2} z$.

5. Essentially bounded functions

Suppose that $(X, \mathcal{A}, \mu)$ is a measure space with $\mu(X) = 1$. Then Hölder's inequality shows that for f measurable, $\|f\|_{L^p(\mu)} \in [0, \infty]$ is a nondecreasing function of $p \in (0, \infty)$. Indeed, if $0 < p_1 < p_2 < \infty$, $\|f\|_{L^{p_2}(\mu)} < \infty$ and $p = \frac{p_2}{p_1}$, then $1 < p < \infty$ and it follows that

$$\begin{aligned} \|f\|_{L^{p_1}(\mu)} &= \left(\int_X |f(x)|^{p_1} \, d\mu(x)\right)^{\frac{1}{p_1}} \leq \left(\int_X |f(x)|^{p_1 p} \, d\mu(x)\right)^{\frac{1}{pp_1}} \left(\int_X d\mu(x)\right)^{\frac{1}{p' p_1}} \\ &\leq \left(\int_X |f(x)|^{p_2} \, d\mu(x)\right)^{\frac{1}{p_2}} = \|f\|_{L^{p_2}(\mu)}. \end{aligned}$$

Thus

$$\|f\|_{\maltese} \equiv \lim_{p\to\infty} \|f\|_{L^p(\mu)} = \sup_{0<p<\infty} \|f\|_{L^p(\mu)} \in [0, \infty].$$

The question now arises as to what $\|f\|_{\maltese}$ actually measures. The answer lies in the following two observations. If $\lambda > \|f\|_{\maltese}$, then

$$\lambda \, |\{|f| > \lambda\}|_\mu^{\frac{1}{p}} \leq \left(\int_{\{|f|>\lambda\}} |f|^p \, d\mu\right)^{\frac{1}{p}} \leq \|f\|_{\maltese},$$

which implies

$$|\{|f| > \lambda\}|_\mu \leq \limsup_{p\to\infty} \left(\frac{\|f\|_{\maltese}}{\lambda}\right)^p = 0.$$

Conversely, if $|\{|f|>\lambda\}|_\mu = 0$, then

$$\|f\|_{L^p(\mu)} = \left(\int_{\{|f|\le\lambda\}} |f|^p \, d\mu + \int_{\{|f|>\lambda\}} |f|^p \, d\mu\right)^{\frac{1}{p}} \le \lambda \mu(X)^{\frac{1}{p}} = \lambda,$$

which implies $\lambda \ge \|f\|_{\maltese}$. Thus we conclude that

$$\|f\|_{\maltese} = \inf\left\{\lambda > 0 : |\{|f|>\lambda\}|_\mu = 0\right\},$$

which suggests we define the *essential supremum* of a measurable function in the following way.

Definition 20. *Suppose that $(X, \mathcal{A}, \mu)$ is a measure space and that $f : X \to \mathbb{C}$ is measurable. The* essential supremum *of f is defined to be*

$$\|f\|_\infty \equiv \inf\left\{\lambda > 0 : |\{|f|>\lambda\}|_\mu = 0\right\}.$$

We set

$$L^\infty(\mu) \equiv \{f \text{ measurable} : \|f\|_\infty < \infty\}.$$

It is easy to show that $L^\infty(\mu)$ is a linear space and that $\|f\|_{L^\infty(\mu)} \equiv \|f\|_\infty$ defines a norm on $L^\infty(\mu)$ (after identifying functions that agree outside a set of measure zero). It is *surprisingly* easy to show that $L^\infty(\mu)$ is complete. Indeed, if $\{f_n\}_{n=1}^\infty$ is a Cauchy sequence in the metric space $L^\infty(\mu)$, then $\{f_n\}_{n=1}^\infty$ converges *uniformly* outside the exceptional set

$$E \equiv \bigcup_{m,n=1}^\infty E_{m,n} \equiv \bigcup_{m,n=1}^\infty \{x \in X : |(f_m - f_n)(x)| > \|f_m - f_n\|_\infty\},$$

to a measurable function $f : (X \setminus E) \to \mathbb{C}$. Since

$$\mu(E) \le \sum_{m,n=1}^\infty \mu(E_{m,n}) = \sum_{m,n=1}^\infty 0 = 0,$$

we may view f as belonging to $L^\infty(\mu)$. It is now evident that $\|f - f_n\|_\infty = \lim_{m\to\infty} \|f_m - f_n\|_\infty$ tends to 0 as $n \to \infty$.

We have already established the first assertion in the following exercise.

Exercise 5. *Suppose that $\mu(X) = 1$ and $\|f\|_{L^r(\mu)} < \infty$ for some $0 < r < \infty$. Then*

(1) $\lim_{p\to\infty} \|f\|_{L^p(\mu)} = \|f\|_\infty$,
(2) $\lim_{p\to 0} \|f\|_{L^p(\mu)} = \exp\left\{\int_X \ln|f| \, d\mu\right\}$.

CHAPTER 8

Complex measures and the Radon-Nikodym theorem

We now wish to extend the notion of a positive measure to complex-valued functionals. We begin with an example.

Example 6. *Given a positive measure ν on a measurable space $(X, \mathcal{A})$, and a complex-valued function $h \in L^1(\nu)$, we can define a set functional μ by*

$$\mu(E) = \int_E h d\nu, \qquad E \in \mathcal{A}.$$

It is easy to verify that μ is a complex measure on $\mathcal{A}$, i.e. that the countable additivity in (0.1) below holds. Indeed, Corollary 14 shows that if $E = \dot{\bigcup}_{n=1}^{\infty} E_n$, then

$$\int_E |h| \, d\nu = \sum_{n=1}^{\infty} \int_{E_n} |h| \, d\nu,$$

and it follows that

$$\left| \int_{\dot{\bigcup}_{n=N+1}^{\infty} E_n} h d\nu \right| \leq \int_{\dot{\bigcup}_{n=N+1}^{\infty} E_n} |h| \, d\nu = \int \sum_{n=N+1}^{\infty} \chi_{E_n} |h| \, d\nu = \sum_{n=N+1}^{\infty} \int_{E_n} |h| \, d\nu \to 0$$

as $N \to \infty$. Now for each $N \geq 1$ we have

$$\begin{aligned} \sum_{n=1}^{N} \mu(E_n) &= \sum_{n=1}^{N} \int_{E_n} h d\nu = \int_{\dot{\bigcup}_{n=1}^{N} E_n} h d\nu \\ &= \int_{E \setminus \left(\dot{\bigcup}_{n=N+1}^{\infty} E_n \right)} h d\nu \\ &= \int_E h d\nu - \int_{\dot{\bigcup}_{n=N+1}^{\infty} E_n} h d\nu, \end{aligned}$$

and taking limits as $N \to \infty$, we get $\sum_{n=1}^{\infty} \mu(E_n) = \int_E h d\nu = \mu(E)$.

More generally, we have the following definition. Consider a measurable space $(X, \mathcal{A})$ and a functional

$$\mu : \mathcal{A} \to \mathbb{C}.$$

Note that we do *not* permit μ to take on infinite values here.

DEFINITION 21. *We say that* μ *is a complex measure on* $\mathcal{A}$, *or on* X, *if for every sequence* $\{E_n\}_{n=1}^{\infty}$ *of pairwise disjoint measurable sets, the series* $\sum_{n=1}^{\infty}\mu(E_n)$ *converges and we have*

$$\mu\left(\dot{\bigcup}_{n=1}^{\infty} E_n\right)=\sum_{n=1}^{\infty}\mu(E_n)\,. \tag{0.1}$$

1. The total variation of a complex measure

The first observation we make is that the convergence of the series in (0.1) must be *absolute*, i.e. $\sum_{n=1}^{\infty}|\mu(E_n)|<\infty$. Indeed, for $0\le k\le 2$, let

$$\mathcal{S}_k=\left\{re^{i\theta}\in\mathbb{C}:0<r<\infty \text{ and } -\frac{\pi}{3}\le\theta-\frac{2\pi k}{3}<\frac{\pi}{3}\right\}$$

denote the sector of aperture $\frac{2\pi}{3}$ centred at the angle $\frac{2\pi k}{3}$. Then with $A_k=\{n:\mu(E_n)\in\mathcal{S}_k\}$ we have

$$\sum_{n=1}^{\infty}|\mu(E_n)|=\sum_{k=0}^{2}\sum_{n\in A_k}|\mu(E_n)|\,,$$

and so if $\sum_{n=1}^{\infty}|\mu(E_n)|=\infty$, there is k such that

$$\sum_{n\in A_k}|\mu(E_n)|=\infty.$$

Without loss of generality we take $k=0$ and note that for $z\in\mathcal{S}_0$ we have

$$\frac{1}{2}|z|\le\operatorname{Re}z\le|z|\,.$$

Thus we conclude that

$$\infty=\sum_{n\in A_0}\operatorname{Re}(\mu(E_n))=\operatorname{Re}\left(\sum_{n\in A_0}\mu(E_n)\right),$$

and so the series $\sum_{n\in A_0}\mu(E_n)$ does *not* converge, contradicting (0.1) and the fact that the sets $\{E_n\}_{n\in A_0}$ are measurable and pairwise disjoint.

The above observation suggests the possibility that there exists a closely related *positive* measure associated with μ, namely the nonnegative set functional $|\mu|:\mathcal{A}\to[0,\infty)$ defined by

$$|\mu|(E)\equiv\sup\left\{\sum_{n=1}^{\infty}|\mu(E_n)|:E=\dot{\bigcup}_{n=1}^{\infty}E_n \text{ with } E_n\in\mathcal{A} \text{ for all } n\ge 1\right\}.$$

This set functional $|\mu|$ is referred to as the *total variation* of μ, and turns out to be a positive measure on $\mathcal{A}$ with $|\mu|(X)<\infty$.

THEOREM 52. *Let* $(X,\mathcal{A})$ *be a measurable space, and suppose* μ *is a complex measure on* $\mathcal{A}$. *Then the total variation* $|\mu|$ *of* μ *is a positive measure on* $\mathcal{A}$ *with* $|\mu|(X)<\infty$.

Proof: To prove the inequality

$$|\mu|(E)\le\sum_{n=1}^{\infty}|\mu(E_n)|\,, \tag{1.1}$$

let $E = \dot{\bigcup}_{m=1}^{\infty} A_m$. Then

$$\begin{aligned}\sum_{m=1}^{\infty} |\mu(A_m)| &= \sum_{m=1}^{\infty} \left|\sum_{n=1}^{\infty} \mu(A_m \cap E_n)\right| \leq \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} |\mu(A_m \cap E_n)| \\ &= \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |\mu(A_m \cap E_n)| \leq \sum_{n=1}^{\infty} |\mu|(E_n),\end{aligned}$$

and if we take the supremum over all decompositions $E = \dot{\bigcup}_{m=1}^{\infty} A_m$ we obtain (1.1).

Now we turn to proving

$$|\mu|(E) \geq \sum_{n=1}^{\infty} |\mu(E_n)|. \tag{1.2}$$

Since at this point $|\mu(E_n)|$ could be infinite, we cannot use $|\mu(E_n)| - \varepsilon < |\mu(E_n)|$ for a small positive ε, and instead we let t_n be *any* nonegative real number satisfying $t_n < |\mu(E_n)|$. Then there is a decomposition $E_n = \dot{\bigcup}_{m=1}^{\infty} A_m^n$ satisfying

$$t_n < \sum_{m=1}^{\infty} |\mu(A_m^n)|. \tag{1.3}$$

It follows that

$$\sum_{n=1}^{\infty} t_n \leq \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} |\mu(A_m^n)| \leq |\mu|(E),$$

and taking the supremum over sequences $\{t_n\}_{n=1}^{\infty}$ satisfying (1.3), we obtain (1.2).

Finally, we prove that $|\mu|(X) < \infty$. Suppose, in order to derive a contradiction, that $|\mu|(X) = \infty$. Then there is a decomposition $X = \dot{\bigcup}_{n=1}^{\infty} E_n$ with

$$\sum_{n=1}^{\infty} |\mu(E_n)| > 6(|\mu(X)| + 1).$$

Using the notation introduced before the statement of the Theorem 52 we have

$$6(|\mu(X)| + 1) < \sum_{n=1}^{\infty} |\mu(E_n)| = \sum_{k=0}^{2} \left\{\sum_{n \in A_k} |\mu(E_n)|\right\},$$

and so there is $k \in \{0, 1, 2\}$ such that $\sum_{n \in A_k} |\mu(E_n)| > 2(|\mu(X)| + 1)$. Without loss of generality, $k = 0$ and using $\frac{1}{2}|z| \leq \operatorname{Re} z \leq |z|$ for $z \in \mathcal{S}_0$ we have

$$2(|\mu(X)| + 1) < \sum_{n \in A_0} |\mu(E_n)| \leq 2 \sum_{n \in A_0} \operatorname{Re}(\mu(E_n)) = 2 \operatorname{Re}\left(\sum_{n \in A_0} \mu(E_n)\right),$$

and thus

$$\left|\sum_{n \in A_0} \mu(E_n)\right| \geq \operatorname{Re}\left(\sum_{n \in A_0} \mu(E_n)\right) > |\mu(X)| + 1.$$

Now set $A = \bigcup_{n \in A_0} E_n$ and $B = X \setminus A$ so that

$$\begin{aligned} |\mu(A)| &= \left| \sum_{n \in A_0} \mu(E_n) \right| > |\mu(X)| + 1 > 1, \\ |\mu(B)| &= |\mu(X) - \mu(A)| \geq |\mu(A)| - |\mu(X)| \\ &> |\mu(X)| + 1 - |\mu(X)| = 1. \end{aligned}$$

Now $\infty = |\mu|(X) = |\mu|(A) + |\mu|(B)$ implies that at least one of A and B has infinite $|\mu|$-measure, say B. Then we define $A_1 = A$ and $B_1 = B$ so that

$$X = A_1 \dot{\bigcup} B_1 \text{ with } |\mu(A_1)| > 1 \text{ and } |\mu|(B_1) = \infty.$$

Now iterate this construction with B_1 in place of X to obtain measurable sets A_2 and B_2 such that

$$B_1 = A_2 \dot{\bigcup} B_2 \text{ with } |\mu(A_2)| > 1 \text{ and } |\mu|(B_2) = \infty.$$

Continuing by induction we obtain sequences $\{A_n\}_{n=1}^{\infty}$ and $\{B_n\}_{n=1}^{\infty}$ of measurable sets satisfying

$$B_{n-1} = A_n \dot{\bigcup} B_n \text{ with } |\mu(A_n)| > 1 \text{ and } |\mu|(B_n) = \infty, \qquad n \geq 2.$$

Now let $A = \dot{\bigcup}_{n=1}^{\infty} A_n$ be the union of the pairwise disjoint sets $\{A_n\}_{n=1}^{\infty}$. Then we must have

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A_n),$$

but this is impossible since the series on the right is divergent: $|\mu(A_n)| > 1$ for all n.

Definition 22. *Let $(X, \mathcal{A})$ be a measurable space. If μ, ν are complex measures on X, then so is $\alpha\mu + \beta\nu$ for $\alpha, \beta \in \mathbb{C}$ where*

$$(\alpha\mu + \beta\nu)(E) = \alpha\mu(E) + \beta\nu(E), \qquad E \in \mathcal{A}.$$

Denote by $\mathsf{M}(X)$ the normed linear space of complex measures on X with norm given by

$$\|\mu\| \equiv |\mu|(X), \qquad \mu \in \mathsf{M}(X).$$

Exercise 6. *Show that $\mathsf{M}(X)$ is complete, hence a Banach space.*

2. The Radon-Nikodym theorem

Every complex number z has a representation in polar coordinates as $z = \zeta |z|$ where $|z| \geq 0$ and $|\zeta| = 1$ (we usually write $\zeta = e^{i\theta}$ as well). It turns out that there is a similar representation of a complex measure μ on a measurable space $(X, \mathcal{A})$ as (see Example 6 above)

$$\mu = \zeta |\mu|, \tag{2.1}$$

where $|\mu|$ is the total variation of μ, and ζ is a measurable function on X satisfying $|\zeta(x)| = 1$ for all $x \in X$. This representation of a complex measure μ is often called the *polar representation* of μ.

In the special case that μ takes on only *real* values, we call μ a real measure, and in the polar representation (2.1), we have $\zeta(x) = \pm 1$ for all $x \in X$. In particular, if

$$\mu_1 = \chi_{\{x \in X : \zeta(x) = 1\}} \mu \text{ and } \mu_2 = -\chi_{\{x \in X : \zeta(x) = -1\}} \mu,$$

then both μ_1 and μ_2 are *positive* measures on X whose difference $\mu_1 - \mu_2$ is μ, and μ_1 and μ_2 are carried by *disjoint* sets. We say that a positive or complex measure μ on $\mathcal{A}$ is *carried* by a set A if $\mu(E) = 0$ for all $E \in \mathcal{A}$ such that $A \cap E = \emptyset$.

This decomposition $\mu = \mu_1 - \mu_2$, where the μ_i are positive measures carried by disjoint sets, is called the *Hahn decomposition* of the real measure μ. Note also that $|\mu| = \mu_1 + \mu_2$. A much simpler decomposition is the *Jordan decomposition*

$$\mu = \frac{1}{2}(|\mu| + \mu) - \frac{1}{2}(|\mu| - \mu) \equiv \mu_+ - \mu_-, \tag{2.2}$$

where $\mu_\pm$ are easily shown to be positive measures, but no claim is made regarding $\mu_\pm$ being carried by disjoint sets. It turns out that $\mu_1 = \mu_+$ and $\mu_2 = \mu_-$ so that $\mu_\pm$ are indeed carried by disjoint sets. But this is hard to prove, and we will obtain it from a much more general, and significantly deeper, decomposition of a complex measure; namely the Radon-Nikodym Theorem. To state this most important of the theorems in measure theory, we need some definitions.

Definition 23. *Let $(X, \mathcal{A})$ be a measurable space. Suppose that μ, ν are measures (complex or positive) on $\mathcal{A}$ and that λ is a positive measure on $\mathcal{A}$. Then*

1. *μ is said to be* concentrated on *(or* carried by *or* lives on*) a measurable set $A \in \mathcal{A}$ if*
$$\mu(E) = \mu(E \cap A) \text{ for all } E \in \mathcal{A},$$
equivalently,
$$\mu(E) = 0 \text{ for all } E \in \mathcal{A} \text{ with } E \cap A = \emptyset;$$
2. *μ and ν are said to be* mutually singular *if there are disjoint measurable sets $A, B \in \mathcal{A}$ such that μ is concentrated on A and ν is concentrated on B. In this case we write $\mu \perp \nu$;*
3. *μ is said to be* absolutely continuous *with respect to the* positive *measure λ if*
$$\mu(E) = 0 \text{ for all null sets } E \text{ of } \lambda.$$
In this case we write $\mu \ll \lambda$.

Note that if in the first definition, the measure μ is a positive measure, then μ is concentrated on a set A if and only if $\mu(A^c) = 0$. Of course this simple characterization *doesn't* extend to complex measures μ. The following properties of these definitions are easy to prove.

Proposition 9. *Let $(X, \mathcal{A})$ be a measurable space. Suppose that μ, ν are measures (complex or positive) on $\mathcal{A}$ and that λ is a positive measure on $\mathcal{A}$. Then*

1. *the connections with the total variation of a measure are these:*
   (a) *μ is concentrated on A if and only if $|\mu|$ is concentrated on A;*
   (b) *$\mu \perp \nu$ if and only if $|\mu| \perp |\nu|$;*
   (c) *$\mu \ll \lambda$ if and only if $|\mu| \ll \lambda$;*
2. *the connections with the additive structure of measures are these:*
   (a) *$\mu \perp \lambda$ and $\nu \perp \lambda \Longrightarrow (\mu + \nu) \perp \lambda$.*
   (b) *$\mu \ll \lambda$ and $\nu \ll \lambda \Longrightarrow (\mu + \nu) \ll \lambda$.*

(3) *the connections between* $\ll$ *and* $\perp$ *are these:*
   (a) $\mu \ll \lambda$ *and* $\nu \perp \lambda \Longrightarrow \mu \perp \nu$.
   (b) $\mu \ll \lambda$ *and* $\mu \perp \lambda \Longrightarrow \mu = 0$.

EXERCISE 7. *Prove this proposition.*

DEFINITION 24. *A positive measure* λ *on a measurable space* $(X, \mathcal{A})$ *is* σ*-finite if* $X = \bigcup_{n=1}^{\infty} X_n$ *is a countable union of measurable sets* X_n *with* $\lambda(X_n) < \infty$.

Now we can state the most important of the theorems in measure theory. It gives, under certain conditions, a decomposition of a complex measure μ into one piece that is *absolutely continous* with respect to a given positive measure λ, and another piece that is *mutually singular* with respect to λ. Moreover, it describes completely the nature of the absolutely continuous piece, and shows that the measures in Example 6 are the only such pieces!

THEOREM 53 (Radon-Nikodym Theorem). *Let* $(X, \mathcal{A})$ *be a measurable space. Suppose that* $\mu \in \mathsf{M}(X)$ *is a complex measure and that* λ *is a positive* σ*-finite measure on* $\mathcal{A}$.

(1) *There is a unique pair of complex measures* $\mu_a, \mu_s \in \mathsf{M}(X)$ *such that*

$$\mu = \mu_a + \mu_s \text{ where } \mu_a \ll \lambda \text{ and } \mu_s \perp \lambda.$$

If in addition μ *is positive (and thus finite), so are* μ_a *and* μ_s.

(2) *There is a unique* $h \in L^1(\lambda)$ *such that*

$$\mu_a(E) = \int_E h d\lambda, \qquad \text{for all } E \in \mathcal{A}. \tag{2.3}$$

The function $h \in L^1(\lambda)$ in part (2) of the theorem is called the Radon-Nikodym derivative of μ with respect to λ and is usually denoted

$$h = \frac{d\mu}{d\lambda}.$$

This function will be obtained using the Riesz representation theorem 51 for an associated Hilbert space $L^2(\varphi)$.

Proof: We begin with the proof of uniqueness. If

$$\mu_a + \mu_s = \mu'_a + \mu'_s$$

where $\mu_a, \mu'_a \ll \lambda$ and $\mu_s, \mu'_s \perp \lambda$ then

$$\omega \equiv \mu_a - \mu'_a = \mu'_s - \mu_s \text{ satisfies } \omega \ll \lambda \text{ and } \omega \perp \lambda,$$

hence by Proposition 9 (3)(b) we have $\omega = 0$. The uniqueness of $h \in L^1(\lambda)$ in part (2) is simply the fact that $\int_E h d\lambda = 0$ for all $E \in \mathcal{A}$ implies $h = 0$ λ-almost everywhere.

Conversely, we first prove the special case where both μ and λ are positive finite measures. Then the sum $\varphi = \mu + \lambda$ is also a positive finite measure, and the Cauchy-Schwarz inequality gives

$$\left| \int f d\mu \right| \leq \int |f| \, d\mu \leq \int |f| \, d\varphi \leq \left(\int |f|^2 \, d\varphi \right)^{\frac{1}{2}} \sqrt{\varphi(X)} = \sqrt{\varphi(X)} \, \|f\|_{L^2(\varphi)}$$

for every $f \in L^2(\varphi)$. Thus we see that the map

$$\Lambda f \equiv \int f d\mu, \qquad f \in L^2(\varphi),$$

defines a bounded linear functional on the Hilbert space $L^2(\varphi)$! By the Riesz representation theorem 51 for Hilbert spaces, there is a unique $g \in L^2(\varphi)$ such that

$$\int fgd\varphi = \langle f, \overline{g} \rangle_{L^2(\varphi)} = \Lambda f = \int fd\mu, \qquad f \in L^2(\varphi).$$

We now claim that $g(x) \in [0,1]$ for φ-almost every $x \in X$. To see this, consider a ball $B(z,r)$ in the complex plane that doesn't intersect $[0,1]$, i.e. $B(z,r) \cap [0,1] = \emptyset$. Let $E = g^{-1}(B(z,r))$ and assume, in order to derive a contradiction, that $\varphi(E) > 0$. Then we have

$$\begin{aligned} \frac{\mu(E)}{\varphi(E)} &= \frac{1}{\varphi(E)} \int \chi_E d\mu = \frac{1}{\varphi(E)} \Lambda \chi_E = \frac{1}{\varphi(E)} \int \chi_E g d\varphi \\ &= \frac{1}{\varphi(E)} \int \chi_E z d\varphi + \frac{1}{\varphi(E)} \int \chi_E (g - z) d\varphi \\ &= z + \frac{1}{\varphi(E)} \int \chi_E (g - z) d\varphi, \end{aligned}$$

which shows that

$$\left| \frac{\mu(E)}{\varphi(E)} - z \right| \leq \frac{1}{\varphi(E)} \int \chi_E |g - z| d\varphi < \frac{1}{\varphi(E)} \int \chi_E r d\varphi = r,$$

since $|g(x) - z| < r$ for $x \in E$. Thus we have shown that $\frac{\mu(E)}{\varphi(E)} \in B(z,r)$. Since $\frac{\mu(E)}{\varphi(E)} \in [0,1]$, we have the desired contradiction to $B(z,r) \cap [0,1] = \emptyset$.

Now let $\{B(z_n, r_n)\}_{n=1}^{\infty}$ be a countable collection of balls satisfying

$$\mathbb{C} \setminus [0,1] = \bigcup_{n=1}^{\infty} B(z_n, r_n).$$

It follows that

$$\mu\left(g^{-1}(\mathbb{C} \setminus [0,1])\right) \leq \sum_{n=1}^{\infty} \mu\left(g^{-1}(B(z_n, r_n))\right) = \sum_{n=1}^{\infty} 0 = 0,$$

which says that $g(x) \in [0,1]$ for φ-almost every $x \in X$.

Thus we may assume that $g(x) \in [0,1]$ for *all* $x \in X$. We then have

$$\begin{aligned} \int_X (1-g) fd\mu &= \int_X fd\mu - \int_X fgd\mu = \int fgd\varphi - \int_X fgd\mu \\ &= \int_X fgd(\mu + \lambda) - \int_X fgd\mu = \int_X fgd\lambda, \end{aligned} \tag{2.4}$$

for all $f \in L^2(\varphi)$. We can now define μ_a and μ_s. Formally, we expect that

$$(1-g) fd\mu = fgd\lambda, \text{ hence } \frac{d\mu}{d\lambda} = \frac{g}{1-g},$$

and this suggests that μ_a should live where $g < 1$ and that μ_s should live where $g = 1$. So let

$$\begin{aligned} A &\equiv \{x \in X : 0 \leq g(x) < 1\}, \\ S &\equiv \{x \in X : g(x) = 1\}, \end{aligned}$$

and set

$$\begin{aligned} \mu_a(E) &\equiv \mu(E \cap A), \qquad E \in \mathcal{A}, \\ \mu_s(E) &\equiv \mu(E \cap S), \qquad E \in \mathcal{A}. \end{aligned}$$

It is easy to see that $\mu_s \perp \lambda$. Indeed, since $g = 1$ on S we have

$$\lambda(S) = \int \chi_S d\lambda = \int \chi_S g d\lambda = \int (1-g)\chi_S d\mu = 0,$$

which means that λ is concentrated on $A = S^c$, while by definition μ_s is concentrated on S. To see that $\mu_a \ll \lambda$ is not much harder. If $E \in \mathcal{A}$ satisfies $\lambda(E) = 0$, then with $f = \chi_{E \cap A}$ in (2.4) we have from (2.4) that

$$0 = \int_{E \cap A} g d\lambda = \int_X f g d\lambda = \int_X (1-g) f d\mu = \int_{E \cap A} (1-g) d\mu.$$

Since $1 - g > 0$ on A we conclude that $\mu(E \cap A) = 0$, i.e. $\mu_a(E) = 0$.

Finally, to see that there is $h \in L^1(\lambda)$ satisfying (2.3), we note that for $E \in \mathcal{A}$ and $n \geq 1$, equation (2.4) with $f = (1 + g + ... + g^n)\chi_E$ and the Monotone Convergence Theorem applied twice yields

$$\begin{aligned} \mu(E \cap A) &= \int_E \left\{ \lim_{n \to \infty} \left(1 - g^{n+1}\right) \right\} d\mu \\ &= \lim_{n \to \infty} \int_E \left(1 - g^{n+1}\right) d\mu = \lim_{n \to \infty} \int_X (1-g) \left\{(1 + g + ... + g^n)\chi_E\right\} d\mu \\ &= \lim_{n \to \infty} \int_X \left\{(1 + g + ... + g^n)\chi_E\right\} g d\lambda \\ &= \lim_{n \to \infty} \int_E \left(g + g^2 + ... + g^{n+1}\right) d\lambda \\ &= \int_E \frac{g}{1-g} d\lambda, \end{aligned}$$

since both

$$\left(1 - g^{n+1}\right) \nearrow 1 \text{ and } \left(g + g^2 + ... + g^{n+1}\right) \nearrow \frac{g}{1-g}$$

pointwise as $n \to \infty$. Thus $h = \frac{g}{1-g} \in L^1(\lambda)$ and (2.3) holds. Note that both μ_a and μ_s are positive measures, and that h is nonnegative.

Now we remove the additional assumptions on μ and λ. First, we consider the case where μ is positive and finite, and λ is positive and σ-finite. It is easy to construct a pairwise disjoint decomposition $X = \dot{\bigcup}_{n=1}^{N \text{ or } \infty} X_n$ such that $0 < \lambda(X_n) < \infty$ for all n. Then define

$$w \equiv \sum_{n=1}^{N \text{ or } \infty} \frac{1}{2^n (1 + \lambda(X_n))} \chi_{X_n}.$$

Then we have both

$$0 < w(x) < 1 \text{ for all } x \in X,$$

and

$$0 < \int_X w d\lambda = \sum_{n=1}^{N \text{ or } \infty} \frac{\lambda(X_n)}{2^n (1 + \lambda(X_n))} < 1.$$

If we let λ_0 be the finite positive measure given by

$$\lambda_0(E) = \int_E w d\lambda, \qquad E \in \mathcal{A},$$

then from what we have already proved we obtain

$$\begin{aligned} \mu &= \mu_a + \mu_s, \qquad \text{where } \mu_a \ll \lambda_0 \text{ and } \mu_s \perp \lambda_0, \\ \mu_a(E) &= \int_E h d\lambda_0, \qquad \text{for all } E \in \mathcal{A}, \text{ where } 0 \le h \in L^1(\lambda_0). \end{aligned}$$

Clearly $\mu_a \ll \lambda$ and $\mu_s \perp \lambda$ both hold, as well as

$$\mu_a(E) = \int_E h_0 d\lambda_0 = \int_E h_0 w d\lambda, \qquad \text{for all } E \in \mathcal{A},$$

so that $h = h_0 w \in L^1(\lambda)$ satisfies (2.3). Indeed, $\int_E f d\lambda_0 = \int_E f w d\lambda$ for all $f = \chi_F$ with $F \in \mathcal{A}$, hence for all f simple, hence for all f nonnegative including $f = h_0$.

Finally, to remove the restriction that μ is positive, write $\mu = \nu_+ - \nu_- + i\omega_3 - i\omega_4$ where ν and ω are the real and imaginary parts of μ and $\nu = \nu_+ - \nu_-$ and $\omega = \omega_+ - \omega_-$ are the Jordan decompositions of ν and ω respectively as defined in (2.2).

REMARK 25. *The σ-finiteness of λ cannot be dropped from the hypotheses of the Radon-Nikodym theorem. For example, if μ is Lebesgue measure on $[0,1]$ and λ is counting measure on $[0,1]$, then $\mu \ll \lambda$ but if there were $h \in L^1(\lambda)$ such that $\mu(E) = \int_E h d\lambda$, then we'd have $h(x) = \int_{\{x\}} h d\lambda = \mu(\{x\}) = 0$ for all $x \in [0,1]$, yielding the contradiction $\mu = 0$.*

We can now obtain as corollaries, both the *polar representation* of a complex measure and the *Hahn decomposition* of a real measure. We note that $\mu \ll |\mu|$ holds trivially where $|\mu|$ is the total variation of μ.

COROLLARY 18 (Polar representation). *Let μ be a complex measure on $(X, \mathcal{A})$. Then the Radon-Nikodym derivative $h = \frac{d\mu}{d|\mu|}$ satisfies $|h(x)| = 1$ for $|\mu|$-almost every $x \in X$.*

Thus for a complex measure, we can write $d\mu(x) = e^{i\theta(x)} d|\mu|(x)$, explaining the term polar representation.

Proof: We first claim that if $0 < r < 1$ and $E_r \equiv \{x \in X : |h(x)| < r\}$, then $|\mu|(E_r) = 0$. Indeed, if $E_r = \dot{\bigcup}_{n=1}^{\infty} F_n$ then

$$\sum_{n=1}^{\infty} |\mu(F_n)| = \sum_{n=1}^{\infty} \left| \int_{F_n} h d|\mu| \right| \le \sum_{n=1}^{\infty} \int_{F_n} |h| d|\mu| \le r \sum_{n=1}^{\infty} |\mu|(F_n) = r|\mu|(E_r).$$

Taking the supremum over all decompostions $E_r = \dot{\bigcup}_{n=1}^{\infty} F_n$ we obtain $0 \le |\mu|(E_r) \le r|\mu|(E_r)$, which implies $|\mu|(E_r) = 0$ since $r < 1$. It follows that

$$|\mu|(\{x \in X : |h(x)| < 1\}) = \lim_{n\to\infty} |\mu|\left(\left\{x \in X : |h(x)| < 1 - \frac{1}{n}\right\}\right) = 0.$$

To show that $|\mu|(\{x \in X : |h(x)| > 1\})$ vanishes, we apply the averaging argument used in the proof of the Radon-Nikodym theorem. It suffices to show that if

$B(z,r)\cap\overline{B(0,1)}=\emptyset$, then the subset $E\equiv h^{-1}(B(z,r))$ of X satisfies $|\mu|(E)=0$. But if $|\mu|(E)>0$, then we obtain

$$\begin{aligned}\frac{\mu(E)}{|\mu|(E)} &= \frac{1}{|\mu|(E)}\int\chi_E d\mu=\frac{1}{|\mu|(E)}\int\chi_E h d\varphi\\ &= \frac{1}{|\mu|(E)}\int\chi_E z d\varphi+\frac{1}{|\mu|(E)}\int\chi_E(h-z)\,d\varphi\\ &= z+\frac{1}{|\mu|(E)}\int\chi_E(g-z)\,d\varphi,\end{aligned}$$

which shows that

$$\left|\frac{\mu(E)}{|\mu|(E)}-z\right|\leq\frac{1}{|\mu|(E)}\int\chi_E|h-z|\,d|\mu|<\frac{1}{|\mu|(E)}\int\chi_E r d|\mu|=r,$$

contradicting $B(z,r)\cap\overline{B(0,1)}=\emptyset$.

COROLLARY 19 (Hahn decomposition). *Let μ be a real measure on $(X,\mathcal{A})$. If $\mu=\mu_+-\mu_-$ is the Jordan decomposition of μ, i.e. $\mu_\pm=\frac{1}{2}(|\mu|\pm\mu)$, then $\mu_+\perp\mu_-$. Moreover, if $h=\frac{d\mu}{d|\mu|}$ is the Radon-Nikodym derivative of μ with respect to its total variation $|\mu|$, then $|h(x)|=1$ for $|\mu|$-almost every $x\in X$, and for $E\in\mathcal{A}$ we have*

$$\begin{aligned}\mu_+(E) &= |\mu|(E\cap\{h=1\}),\\ \mu_-(E) &= |\mu|(E\cap\{h=-1\}).\end{aligned}$$

REMARK 26. *Using the Radon-Nikodym theorem it is easy to see that if μ is a complex measure and λ is a σ-finite positive measure, then $\mu\ll\lambda$* if and only if *for every $\varepsilon>0$ there is $\delta>0$ such that*

$$|\mu(E)|<\varepsilon \text{ whenever } \lambda(E)<\delta. \tag{2.5}$$

Indeed, if $f=\frac{d\mu}{d\lambda}\in L^1(\lambda)$ is the Radon-Nikodym derivative, and if $\varepsilon>0$, the Dominated Convergence Theorem shows that there is $M<\infty$ such that $\int_{\{|f|>M\}}|f|\,d\lambda<\frac{\varepsilon}{2}$. Then with $\delta=\frac{\varepsilon}{2M}>0$, we have

$$\begin{aligned}|\mu(E)| &= \left|\int_E f d\lambda\right|\leq\int_{\{|f|>M\}}|f|\,d\lambda+\int_{E\cap\{|f|\leq M\}}|f|\,d\lambda\\ &< \frac{\varepsilon}{2}+M\lambda(E)<\frac{\varepsilon}{2}+M\delta=\varepsilon,\end{aligned}$$

if $\lambda(E)<\delta$.
In fact, even for general positive measures λ, it is true that $\mu\ll\lambda$ if and only if *(2.5) holds. To see this, suppose there is $\varepsilon>0$ and sets $\{E_n\}_{n=1}^\infty$ with $\lambda(E_n)<\frac{1}{2^n}$ but $|\mu(E_n)|\geq\varepsilon$ for all $n\geq1$. Then*

$$\lambda\left(\bigcup_{n=m}^\infty E_n\right)\leq\sum_{n=m}^\infty 2^{-n}\to0 \text{ as } m\to\infty,$$

and so the set $E=\bigcap_{m=1}^\infty\bigcup_{n=m}^\infty E_n$ yields the desired contradiction to Proposition 9 (1) (c):

$$\begin{aligned}\lambda(E) &= \lim_{m\to\infty}\lambda\left(\bigcup_{n=m}^\infty E_n\right)=0,\\ |\mu|(E) &= \lim_{m\to\infty}|\mu|\left(\bigcup_{n=m}^\infty E_n\right)\geq\liminf_{m\to\infty}|\mu|(E_m)\geq\varepsilon.\end{aligned}$$

CHAPTER 9

Differentiation of integrals

In this chapter we investigate to what extent we can differentiate the Lebesgue integral $\int_{\mathbb{R}^n} f d\lambda_n$ in order to recover the integrand f. In one dimension we have for $f \in C_c(\mathbb{R})$ the two familiar statements of the Fundamental Theorem of Calculus:

$$\begin{aligned} \frac{d}{dx}\int_{-\infty}^{x} f d\lambda &= f(x), \qquad x \in \mathbb{R}, \\ \int_{-\infty}^{\infty} f d\lambda &= \lim_{b\to\infty} F(b) - \lim_{a\to-\infty} F(a), \end{aligned}$$

where F is *any* antiderivative of f. The first of these statements can be rewritten in the equivalent forms

$$\lim_{h\to 0}\frac{1}{h}\int_{x}^{x+h} f d\lambda = \lim_{h\to 0}\frac{\int_{-\infty}^{x+h} f d\lambda - \int_{-\infty}^{x} f d\lambda}{h} = f(x),$$

and

$$\lim_{|I|\to 0:\ x\in I}\frac{1}{|I|}\int_I f d\lambda = f(x),$$

for all $f \in C_c(\mathbb{R})$ and $x \in \mathbb{R}$. The latter limit is taken over all intervals I that contain the point x and the assertion is that for $\varepsilon > 0$ there is $\delta > 0$ such that $\left|\frac{1}{|I|}\int_I f d\lambda - f(x)\right| < \varepsilon$ whenever $x \in I$ and $|I| < \delta$. This suggests the following analogue in higher dimensional Euclidean space $\mathbb{R}^n$.

Problem 4. *To what extent is it true that*

$$\lim_{|I|\to 0:\ x\in I}\frac{1}{|I|}\int_I f d\lambda = f(x) \tag{0.6}$$

for $f \in L^1(\mathbb{R}^n)$, $x \in \mathbb{R}^n$, and a family $\{I\}_{x\in I}$ of subsets of $\mathbb{R}^n$ containing x?

Of course, for *continuous* functions f, the above limit (0.6) holds at every $x \in \mathbb{R}^n$, provided only that the sets I have diameters that shrink to 0 as their measures $|I|$ tend to zero. More generally, we will see that for *integrable* functions f, and for sets I which are sufficiently like balls, the above limit (0.6) holds for almost every x in $\mathbb{R}^n$. The proof follows these lines:

- The limit (0.6) holds for every x if f is continuous.
- The space of continuous functions is dense in $L^1(\mathbb{R}^n)$.
- The oscillation of the limit in (0.6) is near zero except on a small set when $\|f\|_{L^1(\mathbb{R}^n)}$ is small.
- The connection between the oscillation of the limit of averages of f in (0.6), and the $L^1(\mathbb{R}^n)$ norm of f, is governed by the maximal function $\mathcal{M}f$ and a *weak type* inequality.

1. Covering lemmas, maximal functions and differentiation

Let

$$\mathcal{D} = \left\{2^k \left(j + [0,1)^n\right)\right\}_{j\in\mathbb{Z}^n, k\in\mathbb{Z}} \equiv \left\{Q_j^k\right\}_{j\in\mathbb{Z}^n, k\in\mathbb{Z}}$$

be the grid of dyadic cubes in $\mathbb{R}^n$, and define the *dyadic maximal function* $\mathcal{M}^{dy} f$ of a locally integrable function f on $\mathbb{R}^n$ by

$$\mathcal{M}^{dy} f(x) \equiv \sup_{x\in Q\in\mathcal{D}} \frac{1}{|Q|} \int_Q |f(y)|\, dy, \qquad x \in \mathbb{R}^n.$$

We say that f is locally integrable, written $f \in L^1_{loc}(\mathbb{R}^n)$, if $f\chi_{B(0,R)} \in L^1(\mathbb{R}^n)$ for all $R < \infty$. Clearly, $\mathcal{M}^{dy} f$ is measurable since it is the supremum over m of the functions

$$\begin{aligned} f_m(x) &= \sum_{j\in\mathbb{Z}^n} \left(\mathbb{E}_{Q_j^m} |f|\right) \chi_{Q_j^m}(x), \qquad x \in \mathbb{R}^n, \\ \mathbb{E}_Q g &\equiv \frac{1}{|Q|} \int_Q g\, d\lambda_n. \end{aligned}$$

Thus $\mathcal{M}^{dy} f(x)$ is the least upper bound of all the dyadic averages $\mathbb{E}_Q |f|$ of $|f|$ at x. In order to study the convergence of the dyadic averages of f, we consider the limit superior of the dyadic averages of $|f|$ at x:

$$\Gamma^{dy} f(x) = \limsup_{Q\to x} \frac{1}{|Q|} \int_Q |f(y)|\, dy, \qquad x \in \mathbb{R}^n,$$

where it is understood by the expression $Q \to x$ that Q is a dyadic cube containing x whose side length is shrinking to zero in the limit. Clearly we have

$$\begin{aligned} \Gamma^{dy}(f - f(x))(x) = 0 &\implies \lim_{Q\to x} \frac{1}{|Q|} \int_Q |f(y) - f(x)|\, dy = 0 \\ &\implies f(x) = \lim_{Q\to x} \frac{1}{|Q|} \int_Q f(y)\, dy. \end{aligned} \tag{1.1}$$

Of course we have

$$\Gamma^{dy} f(x) \le \mathcal{M}^{dy} f(x), \tag{1.2}$$

and the key properties of the maximal operator $\mathcal{M}^{dy}$ are that it is bounded on $L^\infty(\mathbb{R}^n)$ and of weak type $1-1$ on $L^1(\mathbb{R}^n)$:

$$\left|\left\{x \in \mathbb{R}^n : \mathcal{M}^{dy} f(x) > \lambda\right\}\right| \le \frac{1}{\lambda} \int_{\mathbb{R}^n} |f(y)|\, dy, \qquad \lambda > 0. \tag{1.3}$$

To see (1.3) define

$$\begin{aligned} \Omega_\lambda &= \left\{x \in \mathbb{R}^n : \mathcal{M}^{dy} f(x) > \lambda\right\}, \\ \Phi_\lambda &= \left\{Q \in \mathcal{D} : \frac{1}{|Q|} \int_Q |f(y)|\, dy > \lambda\right\}, \end{aligned}$$

and let $\{Q_m\}_m$ be the set of *maximal* dyadic cubes in Φ_λ. Then the cubes Q_m are pairwise disjoint and we have

$$\Omega_\lambda = \bigcup_{Q\in\Phi_\lambda} Q = \dot{\bigcup_m} Q_m.$$

This is the most successful of covering lemmas: namely we have covered a union Ω_λ of dyadic cubes with a *pairwise disjoint subcollection*. Unravelling definitions yields

$$|\Omega_\lambda| = \sum_m |Q_m| < \sum_m \frac{1}{\lambda} \int_{Q_m} |f(y)|\, dy \le \frac{1}{\lambda} \int_{\mathbb{R}^n} |f(y)|\, dy.$$

The weak type inequality (1.3) for $\mathcal{M}^{dy}$ yields the Lebesgue Differentiation Theorem for dyadic averages.

THEOREM 54. *For $f \in L^1_{loc}(\mathbb{R}^n)$ we have*

$$f(x) = \lim_{|Q| \to 0:\ x \in Q \in \mathcal{D}} \frac{1}{|Q|} \int_Q f(y)\, dy, \qquad a.e. x \in \mathbb{R}^n,$$

in fact,

$$\lim_{|Q| \to 0:\ x \in Q \in \mathcal{D}} \frac{1}{|Q|} \int_Q |f(y) - f(x)|\, dy = 0.$$

Proof: Since the conclusion of the theorem is local it suffices to consider $f \in L^1(\mathbb{R}^n)$ with compact support. Given $\varepsilon > 0$, we can use Lemma 26 to choose $g \in C_c(\mathbb{R}^n)$ with $\int |f - g| < \varepsilon$. However, $\Gamma^{dy}(g - g(x))(x) = 0$ for every $x \in \mathbb{R}^n$ since g is continuous. It follows from the subadditivity of Γ^{dy} and (1.2) that

$$\begin{aligned} \Gamma^{dy}(f - f(x))(x) &\le \Gamma^{dy}(f - f(x) - [g - g(x)])(x) + \Gamma^{dy}(g - g(x))(x) \\ &\le \Gamma^{dy}(f - g)(x) + \Gamma^{dy}(f(x) - g(x))(x) \\ &\le \mathcal{M}^{dy}(f - g)(x) + |(f - g)(x)|. \end{aligned}$$

Now we have

$$\begin{aligned} &\left\{x \in \mathbb{R}^n : \Gamma^{dy}(f - f(x))(x) > \lambda\right\} \\ \subset\ &\left\{x \in \mathbb{R}^n : \mathcal{M}^{dy}(f - g)(x) > \frac{\lambda}{2}\right\} \cup \left\{x \in \mathbb{R}^n : |(f - g)(x)| > \frac{\lambda}{2}\right\} \end{aligned}$$

and so

$$\begin{aligned} \left|\left\{x \in \mathbb{R}^n : \Gamma^{dy}(f - f(x))(x) > \lambda\right\}\right| &\le \left|\left\{x \in \mathbb{R}^n : \mathcal{M}^{dy}(f - g)(x) > \frac{\lambda}{2}\right\}\right| \\ &\quad + \left|\left\{x \in \mathbb{R}^n : |(f - g)(x)| > \frac{\lambda}{2}\right\}\right| \\ &\le \frac{2}{\lambda} \int |f - g| + \frac{2}{\lambda} \int |f - g| < \frac{4}{\lambda} \varepsilon. \end{aligned}$$

Now let $\varepsilon \to 0$ to obtain $\left|\left\{x \in \mathbb{R}^n : \Gamma^{dy}(f - f(x))(x) > \lambda\right\}\right| = 0$ for all $\lambda > 0$. This proves that $\Gamma^{dy}(f - f(x))(x) = 0$ for a.e. $x \in \mathbb{R}^n$, and (1.1) now concludes the proof of Lebesgue's differentiation theorem for dyadic averages.

We now wish to extend Lebesgue's differentiation theorem to more general averages, namely to the collection of *almost-balls* in $\mathbb{R}^n$. Fix a large positive constant C. Then we say that a subset I of $\mathbb{R}^n$ is an *almost-ball* of eccentricity C if there is $r > 0$ and two balls, $B(x, r)$ and $B(y, Cr)$, with

$$B(x, r) \subset I \subset B(y, Cr). \tag{1.4}$$

Note that we do not require x or y to belong to I, nor must x equal y. Thus an almost-ball contains an ordinary ball, and is contained in another ordinary ball of

C times the radius. In order to prove this more general differentiaion theorem, we will use the notion of *shifted dyadic grids* to reduce matters to what we have already proved.

Define a *shifted dyadic grid* to be the collection of cubes

$$\mathcal{D}_\alpha = \left\{2^k \left(j + (-1)^k \alpha + [0,1)^n\right) \ : \ k \in \mathbb{Z}, j \in \mathbb{Z}^n\right\}, \qquad \alpha \in \left\{0, \tfrac{1}{3}, \tfrac{2}{3}\right\}^n . \tag{1.5}$$

The basic properties of these collections are these: In the first place, each $\mathcal{D}_\alpha$ is a grid, namely for $Q, Q' \in \mathcal{D}_\alpha$ we have $Q \cap Q' \in \{\emptyset\, , Q\, , Q'\}$ and Q is a union of 2^n elements of $\mathcal{D}_\alpha$ of equal volume. In the second place, and this is the novel property here, for *any* cube $Q \subset \mathbb{R}^n$, there is a choice of some $\alpha \in \{0, \frac{1}{3}, \frac{2}{3}\}^n$ and some $Q' \in \mathcal{D}_\alpha$ so that

$$Q \subset Q' \text{ and } |Q'| \leq C_n |Q| .$$

Here C_n is a positive constant depending only on dimension n. We prove that $C_1 \leq 4$ in dimension $n = 1$, and leave the general case to the reader. So suppose that $[a, b]$ is an interval. Let $k \in \mathbb{Z}$ be the unique integer satisfying $2^{k-1} < b - a \leq 2^k$. Now choose $j \in \mathbb{Z}$ and $\alpha \in \left\{0, \frac{1}{3}, \frac{2}{3}\right\}$ so that $\left(j + (-1)^{k+1} \alpha\right) 2^{k+1}$ is the largest such expression satisfying $\left(j + (-1)^{k+1} \alpha\right) 2^{k+1} < a$. Then $a \leq \left(j + (-1)^{k+1} \alpha + \frac{1}{3}\right) 2^{k+1}$ and so

$$b \leq 2^k + a \leq 2^k + \left(j + (-1)^{k+1} \alpha + \frac{1}{3}\right) 2^{k+1} \leq \left(j + \frac{5}{6} + (-1)^{k+1} \alpha\right) 2^{k+1}.$$

It follows that

$$[a, b] \subset \left[\left(j + (-1)^{k+1} \alpha\right) 2^{k+1}, \left(j + 1 + (-1)^k \alpha\right) 2^{k+1}\right),$$

where the latter interval belongs to the grid $\mathcal{D}_\alpha$ and has length $2^{k+1} < 4 (b - a)$.

We now define the $\mathcal{D}_\alpha$-analogs of the dyadic maximal operator, namely

$$\mathcal{M}_\alpha^{dy} f(x) = \sup_{Q \in \mathcal{D}_\alpha : \ x \in Q} \frac{1}{|Q|_\mu} \int_Q |f| \ . \tag{1.6}$$

Just as for $\mathcal{M}^{dy}$ we have that $\mathcal{M}_\alpha^{dy}$ is weak type $1 - 1$ on $L^1 (\mathbb{R}^n)$,

$$\left|\left\{x \in \mathbb{R}^n : \mathcal{M}_\alpha^{dy} f (x) > \lambda\right\}\right| \leq \frac{1}{\lambda} \int_{\mathbb{R}^n} |f (y)| \, dy, \qquad \lambda > 0.$$

Now fix $C > 0$ and let $\mathcal{A} = \mathcal{A}_C$ denote the collection of all almost-balls of eccentricity C in $\mathbb{R}^n$. Consider the corresponding maximal function

$$\mathcal{M}^{\mathcal{A}} f (x) \equiv \sup_{I \in \mathcal{A} : \ x \in I} \frac{1}{|I|} \int_I |f (y)| \, dy, \qquad x \in \mathbb{R}^n.$$

For each almost-ball $I \in \mathcal{A}$ and ball $B (y, Cr)$ as in (1.4), there is a cube $Q \supset B (y, Cr)$ with $|Q| \leq (2Cr)^n$. Then the properties of the shifted dyadic grids yield the existence of $\alpha \in \left\{0, \frac{1}{3}, \frac{2}{3}\right\}^n$ and $Q' \in \mathcal{D}_\alpha$ such that

$$I \subset B (y, Cr) \subset Q \subset Q',$$

and

$$|Q'| \leq C_n |Q| \leq C_n (2Cr)^n \leq C_n' |B (x, r)| \leq C_n' |I| .$$

It follows that

$$\frac{1}{|I|} \int_I |f (y)| \, dy \leq \frac{C_n'}{|Q'|} \int_{Q'} |f (y)| \, dy \leq C_n' \mathcal{M}_\alpha^{dy} f (x) , \qquad \text{for each } x \in I,$$

and hence that

$$\mathcal{M}^{\mathcal{A}} f(x) \leq C_n' \max_{\alpha \in \left\{0, \frac{1}{3}, \frac{2}{3}\right\}} \mathcal{M}_\alpha^{dy} f(x).$$

This proves that $\mathcal{M}^{\mathcal{A}}$ is also weak type $1-1$ on $L^1(\mathbb{R}^n)$:

$$\begin{aligned} \left|\left\{x \in \mathbb{R}^n : \mathcal{M}^{\mathcal{A}} f(x) > \lambda\right\}\right| &\leq \sum_{\alpha \in \left\{0, \frac{1}{3}, \frac{2}{3}\right\}} \left|\left\{x \in \mathbb{R}^n : \mathcal{M}_\alpha^{dy} f(x) > \frac{\lambda}{3C_n'}\right\}\right| \\ &\leq \frac{9C_n'}{\lambda} \int_{\mathbb{R}^n} |f(y)|\, dy, \qquad \lambda > 0. \end{aligned}$$

As a result we can prove the following theorem in exactly the same way as Theorem 54 above.

Theorem 55. *Let $\mathcal{A} = \mathcal{A}_C$ be the collection of all almost-balls of eccentricity $C > 0$. For $f \in L^1_{loc}(\mathbb{R}^n)$ we have*

$$f(x) = \lim_{|I| \to 0:\ x \in I \in \mathcal{A}} \frac{1}{|I|} \int_I f(y)\, dy, \qquad a.e.\, x \in \mathbb{R}^n,$$

in fact,

$$\lim_{|I| \to 0:\ x \in I \in \mathcal{A}} \frac{1}{|I|} \int_I |f(y) - f(x)|\, dy = 0, \qquad a.e.\, x \in \mathbb{R}^n.$$

Corollary 20. *Suppose that μ is a complex Borel measure on $\mathbb{R}^n$ and that $\mu \ll \lambda_n$ where λ_n is Lebesgue measure. If $f = \frac{d\mu}{d\lambda_n}$ is the Radon-Nikodym derivative of μ with respect to λ_n, then f can be obtained as a limit of ratios of measures:*

$$f(x) = \lim_{|I| \to 0:\ x \in I \in \mathcal{A}} \frac{\mu(I)}{|I|}, \qquad a.e.\, x \in \mathbb{R}^n.$$

Proof: Apply Theorem 55 using $\mu(I) = \int_I f d\lambda_n = \int_I f(y)\, dy$.

Corollary 21. *Let E be a Lebesgue measurable subset of $\mathbb{R}^n$. Then almost every point in E is densely surrounded by points of E in the sense that*

$$\lim_{r \to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 1 \text{ for almost every } x \in E,$$

while almost every point not in E is densely surrounded by points not in E in the sense that

$$\lim_{r \to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = 0 \text{ for almost every } x \notin E.$$

Proof: Apply Theorem 55 using $I = B(x,r)$ and $f = \chi_E$ so that

$$|E \cap B(x,r)| = \int_{B(x,r)} \chi_E(y)\, dy = \int_I f(y)\, dy.$$

This last corollary gives a surprising insight into the structure of measurable sets, which provides yet another illustration of Littlewood's first principle: measurable sets are almost open sets. Of course it is trivial that *every* point in an open set E is *entirely* surrounded by points of E at a small enough scale.

2. The maximal theorem

Our next theorem will require an expression of the L^p norm of a function f in terms of its distribution function

$$|\{|f| > t\}| = |\{x \in \mathbb{R}^n : |f(x)| > t\}|, \qquad t > 0.$$

We could appeal at this point to the following special case of Fubini's theorem, proved in the next chapter. Suppose that $g : \mathbb{R}^n \to [0, \infty)$ is measurable. Then

$$\begin{aligned} \int_{\mathbb{R}^n} g(x)^p\, dx &= \int_{\mathbb{R}^n} \left\{ \int_0^{g(x)} pt^{p-1} dt \right\} dx \\ &= \int_{\mathbb{R}^n} \left\{ \int_{[0,\infty)} \chi_{\{g>t\}}(x)\, pt^{p-1} dt \right\} dx \\ &= \int_{[0,\infty)} \left\{ \int_{\mathbb{R}^n} \chi_{\{g>t\}}(x)\, pt^{p-1} dx \right\} dt \\ &= \int_{[0,\infty)} |\{g > t\}|\, pt^{p-1} dt. \end{aligned} \tag{2.1}$$

However, we only need the following easy approximation to (2.1):

$$\int_{\mathbb{R}^n} g(x)^p\, dx = \sum_{k=-\infty}^{\infty} \int_{\{2^k < g \le 2^{k+1}\}} g(x)^p\, dx \le 2^p \sum_{k=-\infty}^{\infty} 2^{kp} \left| \left\{ 2^k < g \le 2^{k+1} \right\} \right|. \tag{2.2}$$

Theorem 56. *For $1 < p \le \infty$ we have*

$$\left(\int_{\mathbb{R}^n} \left| \mathcal{M}^{dy} f \right|^p \right)^{\frac{1}{p}} \le C_p \left(\int_{\mathbb{R}^n} |f|^p \right)^{\frac{1}{p}}, \qquad f \in L^p(\mathbb{R}^n).$$

Proof: The following argument is from Marcinkiewicz interpolation. Define $f_\lambda = \chi_{\left\{|f| > \frac{\lambda}{2}\right\}} f$ so that $\mathcal{M}^{dy}(f - f_\lambda) \le \frac{\lambda}{2}$ by the boundedness of $\mathcal{M}^{dy}$ on $L^\infty(\mathbb{R}^n)$: $\left\|\mathcal{M}^{dy} g\right\|_{L^\infty(\mathbb{R}^n)} \le \|g\|_{L^\infty(\mathbb{R}^n)}$. Consequently, by the subadditivity of $\mathcal{M}^{dy}$ we have

$$\mathcal{M}^{dy} f \le \mathcal{M}^{dy}(f - f_\lambda) + \mathcal{M}^{dy} f_\lambda \le \frac{\lambda}{2} + \mathcal{M}^{dy} f_\lambda,$$

and thus

$$\left\{ x \in \mathbb{R}^n : \mathcal{M}^{dy} f(x) > \lambda \right\} \subset \left\{ x \in \mathbb{R}^n : \mathcal{M}^{dy} f_\lambda(x) > \frac{\lambda}{2} \right\}, \tag{2.3}$$

for any $\lambda > 0$. Now use (2.2), (2.3) and then (1.3) applied to f_λ with $\lambda = 2^k$ to obtain

$$\begin{aligned}
\int_{\mathbb{R}^n} \left|\mathcal{M}^{dy} f(x)\right|^p dx &\leq 2^p \sum_{k=-\infty}^{\infty} 2^{kp} \left|\left\{x \in \mathbb{R}^n : \mathcal{M}^{dy} f(x) > \lambda\right\}\right| \\
&\leq 2^p \sum_{k=-\infty}^{\infty} 2^{kp} \left|\left\{x \in \mathbb{R}^n : \mathcal{M}^{dy} f_{2^k}(x) > \frac{2^k}{2}\right\}\right| \\
&\leq 2^p \sum_{k=-\infty}^{\infty} 2^{kp} \left\{\frac{1}{2^{k-1}} \int_{\mathbb{R}^n} \left|f_{2^k}(x)\right| dx\right\} \\
&= 2^{p+1} \sum_{k=-\infty}^{\infty} 2^{k(p-1)} \int_{\{x \in \mathbb{R}^n : |f(x)| > 2^{k-1}\}} |f(x)|\, dx \\
&= 2^{p+1} \int_{\mathbb{R}^n} |f(x)| \left\{\sum_{k:\ 2^k < 2|f(x)|} 2^{k(p-1)}\right\} dx \\
&\leq 2^{2p-1} \frac{1}{1 - 2^{p-1}} \int_{\mathbb{R}^n} |f(x)|^p\, dx,
\end{aligned}$$

since $\sum_{k:\ 2^k < 2|f(x)|} 2^{k(p-1)} < \frac{(2|f(x)|)^{p-1}}{1-2^{p-1}}$. Note that we have used Corollary 9 in order to interchange summation and integration in the penultimate line above.

EXERCISE 8. *Prove the maximal theorem with $\mathcal{M}^{dy}$ replaced by the larger maximal operator $\mathcal{M}^{\mathcal{A}}$.*

2.1. The Haar basis. In our second example of an orthonormal set in Example 5 of Section 3 of Chapter 7, we showed that the collection of Haar functions $\{h_I^\mu\}_{I \in \mathcal{D}}$ is orthonormal in $L^2(\mu)$, but deferred the proof that it is a *basis* until we had Lebesgue's Differentiation Theorem at our disposal. We assumed there that μ is a positive Borel measure on the real line $\mathbb{R}$ satisfying $\mu(I) > 0$ for every $I \in \mathcal{D}$. For convenience we now assume a bit more:

$$\begin{aligned}
\mu(I) &> 0 \text{ for every } I \in \mathcal{D}, \\
\int_0^\infty d\mu &= \int_{-\infty}^0 d\mu = \infty.
\end{aligned} \tag{2.4}$$

We will need the analogues of dyadic differentiation theory for a positive measure μ in place of Lebesgue measure. The following two theorems are proved in exactly the same way as the corresponding results for Lebesgue measure above. For these two theorems we assume that μ is a positive Borel measure on $\mathbb{R}^n$ satisfying $\mu(I) > 0$ for every $I \in \mathcal{D}$, and $\mu(J) = \infty$ for each of the 2^n 'octants' of the form $J = \prod_{i=1}^n J_i$ where J_i is either $(-\infty, 0)$ or $[0, \infty)$.

THEOREM 57. *For $f \in L^1_{loc}(\mathbb{R}^n)$ we have*

$$f(x) = \lim_{|Q| \to 0:\ x \in Q \in \mathcal{D}} \frac{1}{|Q|_\mu} \int_Q f(y)\, d\mu(y), \qquad \mu - a.e. x \in \mathbb{R}^n,$$

in fact,

$$\lim_{|Q| \to 0:\ x \in Q \in \mathcal{D}} \frac{1}{|Q|_\mu} \int_Q |f(y) - f(x)|\, d\mu(y) = 0.$$

DEFINITION 25. *Define the* dyadic μ-maximal function $\mathcal{M}_\mu^{dy} f$ *of a locally μ-integrable function f on $\mathbb{R}^n$ by*

$$\mathcal{M}_\mu^{dy} f(x) \equiv \sup_{x \in Q \in \mathcal{D}} \frac{1}{|Q|_\mu} \int_Q |f(y)| \, d\mu(y), \qquad x \in \mathbb{R}^n.$$

THEOREM 58. *For $1 < p \leq \infty$ we have*

$$\left(\int_{\mathbb{R}^n} \left| \mathcal{M}_\mu^{dy} f \right|^p d\mu \right)^{\frac{1}{p}} \leq C_p \left(\int_{\mathbb{R}^n} |f|^p \, d\mu \right)^{\frac{1}{p}}, \qquad f \in L^p(\mu).$$

Now we return to dimension $n = 1$. Recall that $\mathcal{D} = \bigcup_{k \in \mathbb{Z}} \mathcal{D}_k$ is then the set of dyadic intervals, where $\mathcal{D}_k \equiv \left\{ \left[j2^k, (j+1) 2^k \right) \right\}_{j \in \mathbb{Z}}$. We defined the *Haar function* h_I^μ for $I \in \mathcal{D}$ by

$$h_I^\mu(x) = \sqrt{\frac{\mu(I_-)\mu(I_+)}{\mu(I)}} \left(-\frac{\mathbf{1}_{I_-}(x)}{\mu(I_-)} + \frac{\mathbf{1}_{I_+}(x)}{\mu(I_+)} \right), \qquad x \in \mathbb{R},$$

where I_- and I_+ are the left and right halves of I, referred to as the *children* of I. The collection of Haar functions $\{h_I^\mu\}_{I \in \mathcal{D}}$ was shown to satisfy the elementary properties

$$\mathsf{supp} h_I^\mu \subset I, \qquad \int h_I^\mu d\mu = 0, \quad \int |h_I^\mu|^2 \, d\mu = 1,$$

and most importantly, the crucial orthogonality property,

$$\int h_I^\mu h_J^\mu d\mu = 0, \qquad \text{if } I, J \in \mathcal{D} \text{ and } I \neq J.$$

To see that $\{h_I^\mu\}_{I \in \mathcal{D}}$ is actually an orthonormal *basis* for $L^2(\mu)$, it suffices by Theorem 49 to establish that *Span* $\{h_I^\mu\}_{I \in \mathcal{D}}$ is dense in $L^2(\mu)$. For this we introduce the expectation functions,

$$\mathbb{E}_k^\mu f(x) \equiv \sum_{I \in \mathcal{D}_k} \left\langle f, \frac{1}{|I|_\mu} \mathbf{1}_I \right\rangle_{L^2(\mu)} \mathbf{1}_I(x), \qquad x \in \mathbb{R},\ k \in \mathbb{Z},$$

which for a given $k \in \mathbb{Z}$, are simply the functions that are constant on dyadic intervals I of length 2^k, and where the constant is the μ-average of f on I. We make three elementary observations regarding the functions $\mathbb{E}_k^\mu f$ for $f \in L^2(\mu)$:

$$\begin{aligned} \mathbb{E}_k^\mu f(x) &\to 0 \text{ as } k \to \infty \text{ for every } x \in \mathbb{R}, \\ \mathbb{E}_k^\mu f(x) &\to f(x) \text{ as } k \to -\infty \text{ for } \mu\text{-almost every } x \in \mathbb{R}, \\ |\mathbb{E}_k^\mu f(x)| &\leq \mathcal{M}_\mu^{dy} f(x) \text{ for every } x \in \mathbb{R}, \end{aligned}$$

and the crucial observation,

$$\mathbb{E}_M^\mu f(x) - \mathbb{E}_N^\mu f(x) = \sum_{I \in \mathcal{D}:\ 2^{M+1} \leq |I| \leq 2^N} \langle f, h_I^\mu \rangle_{L^2(\mu)} h_I^\mu(x), \tag{2.5}$$

for all $x \in \mathbb{R}$, and for all integers $M < N$.

The first observation follows from

$$\left| \left\langle f, \frac{1}{|I|_\mu} \mathbf{1}_I \right\rangle_{L^2(\mu)} \right| = \left| \frac{1}{|I|_\mu} \int_I f d\mu \right| \leq \left(\frac{1}{|I|_\mu} \int_I |f|^2 \, d\mu \right)^{\frac{1}{2}} = \frac{1}{\sqrt{|I|_\mu}} \|f\|_{L^2(\mu)}$$

and our second assumption in (2.4). The second observation follows directly from Theorem 57. The third observation is immediate from Definition 25 since

$$\left| \left\langle f, \frac{1}{|I|_\mu} \mathbf{1}_I \right\rangle_{L^2(\mu)} \right| \leq \frac{1}{|I|_\mu} \int_I |f| \, d\mu \leq \mathcal{M}_\mu^{dy} f(x), \qquad \text{for } x \in I.$$

We now turn to the verification of the crucial observation (2.5). It suffices to prove the cases $N = M + 1$, and then add them up. So for $I \in \mathcal{D}$ we must prove that

$$\begin{aligned} &\left\langle f, \frac{1}{|I_-|_\mu} \mathbf{1}_{I_-} \right\rangle_{L^2(\mu)} \mathbf{1}_{I_-}(x) + \left\langle f, \frac{1}{|I_+|_\mu} \mathbf{1}_{I_+} \right\rangle_{L^2(\mu)} \mathbf{1}_{I_+}(x) \\ &\qquad - \left\langle f, \frac{1}{|I|_\mu} \mathbf{1}_I \right\rangle_{L^2(\mu)} \mathbf{1}_I(x) \\ &= \langle f, h_I^\mu \rangle_{L^2(\mu)} h_I^\mu(x), \qquad x \in I, \end{aligned} \tag{2.6}$$

where

$$h_I^\mu(x) = \sqrt{\frac{|I_-|_\mu |I_+|_\mu}{|I|_\mu}} \left(-\frac{\mathbf{1}_{I_-}(x)}{|I_-|_\mu} + \frac{\mathbf{1}_{I_+}(x)}{|I_+|_\mu} \right).$$

This is an elementary but tedious calculation. For $x \in I_-$ the left side of (2.6) is

$$\begin{aligned} &\left\langle f, \frac{1}{|I_-|_\mu} \mathbf{1}_{I_-} \right\rangle_{L^2(\mu)} - \left\langle f, \frac{1}{|I|_\mu} \mathbf{1}_I \right\rangle_{L^2(\mu)} \\ &= \left(\frac{1}{|I_-|_\mu} - \frac{1}{|I|_\mu} \right) \int_{I_-} f d\mu - \frac{1}{|I|_\mu} \int_{I_+} f d\mu \\ &= \left(\frac{|I_+|_\mu}{|I|_\mu} \right) \frac{1}{|I_-|_\mu} \int_{I_-} f d\mu - \frac{1}{|I|_\mu} \int_{I_+} f d\mu, \end{aligned}$$

and the right side is

$$\begin{aligned} &- \langle f, h_I^\mu \rangle_{L^2(\mu)} \sqrt{\frac{|I_-|_\mu |I_+|_\mu}{|I|_\mu}} \frac{1}{|I_-|_\mu} \\ &= -\sqrt{\frac{|I_-|_\mu |I_+|_\mu}{|I|_\mu}} \frac{1}{|I_-|_\mu} \left\{ \sqrt{\frac{|I_-|_\mu |I_+|_\mu}{|I|_\mu}} \left[\int_{I_-} \left(-\frac{1}{|I_-|_\mu} \right) f d\mu + \int_{I_+} \frac{1}{|I_+|_\mu} f d\mu \right] \right\} \\ &= \frac{|I_+|_\mu}{|I|_\mu} \left\{ \frac{1}{|I_-|_\mu} \int_{I_-} f d\mu - \frac{1}{|I_+|_\mu} \int_{I_+} f d\mu \right\}. \end{aligned}$$

Thus (2.6) holds for $x \in I_-$, and the case $x \in I_+$ is similar. This completes the verification of (2.5).

With these observations in hand, we can apply the Dominated Convergence Theorem with umbrella function $g = 3\mathcal{M}_{\mu}^{dy} f$ to obtain

$$\begin{aligned}
(2.7) \qquad & \lim_{M\to-\infty \text{ and } N\to\infty} \left\| f - \sum_{I\in\mathcal{D}:\ 2^{M+1}\leq|I|\leq 2^N} \langle f, h_I^\mu\rangle_{L^2(\mu)} h_I^\mu \right\|_{L^2(\mu)}^2 \\
&= \lim_{M\to-\infty \text{ and } N\to\infty} \int_{\mathbb{R}} \left| f(x) - \sum_{I\in\mathcal{D}:\ 2^{M+1}\leq|I|\leq 2^N} \langle f, h_I^\mu\rangle_{L^2(\mu)} h_I^\mu(x) \right|^2 d\mu(x) \\
&= \int_{\mathbb{R}} \left| f(x) - \lim_{M\to-\infty \text{ and } N\to\infty} \sum_{I\in\mathcal{D}:\ 2^{M+1}\leq|I|\leq 2^N} \langle f, h_I^\mu\rangle_{L^2(\mu)} h_I^\mu(x) \right|^2 d\mu(x) \\
&= \int_{\mathbb{R}} |f(x) - [f(x) - 0]|^2 \, d\mu(x) = 0.
\end{aligned}$$

Note that by Theorem 57 we have $|f(x)| \leq \mathcal{M}_{\mu}^{dy} f(x)$ for μ-almost every $x \in \mathbb{R}$, and so for these x,

$$\begin{aligned}
\left| f(x) - \sum_{I\in\mathcal{D}:\ 2^{M+1}\leq|I|\leq 2^N} \langle f, h_I^\mu\rangle_{L^2(\mu)} h_I^\mu(x) \right| &= \left| f(x) - [\mathbb{E}_M^\mu f(x) - \mathbb{E}_N^\mu f(x)] \right| \\
&\leq |f(x)| + |\mathbb{E}_M^\mu f(x) - \mathbb{E}_N^\mu f(x)| \\
&\leq 3\mathcal{M}_{\mu}^{dy} f(x),
\end{aligned}$$

where $\mathcal{M}_{\mu}^{dy} f \in L^2(\mu)$ by Theorem 58. Thus the umbrella function $g = 3\mathcal{M}_{\mu}^{dy} f$ can be used in the above application of the Dominated Convergence Theorem.

Equation (2.7) shows that *Span* $\{h_I^\mu\}_{I\in\mathcal{D}}$ is dense in $L^2(\mu)$, and Theorem 49 now shows that $\{h_I^\mu\}_{I\in\mathcal{D}}$ is an orthonormal basis for $L^2(\mu)$.

REMARK 27. *We can avoid the use of Theorems 57 and 58 if we appeal to the density of $C_c(\mathbb{R})$ in $L^2(\mu)$. Indeed, we then need only establish (2.7) for $f \in C_c(\mathbb{R})$. This is easy since $\mathbb{E}_k^\mu f(x) \to f(x)$ as $k \to -\infty$ for every $x \in \mathbb{R}$ by continuity of f, and if f is supported in a dyadic interval I, then*

$$\mathcal{M}_{\mu}^{dy} f(x) \leq \|f\|_\infty \mathcal{M}_{\mu}^{dy}(\mathbf{1}_I)(x),$$

and it is easily verified that $\mathcal{M}_{\mu}^{dy}(\mathbf{1}_I) \in L^2(\mu)$. For example, if $I = [0,1)$ then

$$\mathcal{M}_{\mu}^{dy}(\mathbf{1}_I)(x) = \frac{\mu([0,1))}{\mu([0,2^k))}, \qquad 2^{k-1} \leq x < 2^k,\ k \geq 1,$$

and we have

$$\begin{aligned}
& \int_0^\infty \left|\mathcal{M}_\mu^{dy}\left(\mathbf{1}_I\right)(x)\right|^2 d\mu(x) \\
= \quad & \int_0^1 \left|\mathcal{M}_\mu^{dy}\left(\mathbf{1}_I\right)(x)\right|^2 d\mu(x) + \sum_{k=1}^{\infty}\left(\frac{\mu([0,1))}{\mu([0,2^k))}\right)^2 \mu\left(\left[2^{k-1},2^k\right)\right) \\
= \quad & \int_0^1 \left|\mathcal{M}_\mu^{dy}\left(\mathbf{1}_I\right)(x)\right|^2 d\mu(x) + \mu([0,1))^2 \sum_{k=1}^{\infty} \frac{\mu\left(\left[0,2^k\right)\right)-\mu\left(\left[0,2^{k-1}\right)\right)}{\mu([0,2^k))^2} \\
\leq \quad & \mu([0,1)) + \mu([0,1))^2 \sum_{k=1}^{\infty} \int_{\mu([0,2^{k-1}))}^{\mu\left(\left[0,2^k\right)\right)} \frac{1}{t^2}\,dt \\
\leq \quad & \mu([0,1)) + \mu([0,1))^2 \int_{\mu([0,1))}^{\infty} \frac{1}{t^2}\,dt = 2\mu([0,1))\,.
\end{aligned}$$

CHAPTER 10

Product integration and Fubini's theorem

In this chapter we investigate to what extent the order of integration can be *reversed* in a product integral, i.e. when do we have an equality

$$\int_X \left\{ \int_Y f(x,y)\, d\nu(y) \right\} d\mu(x) = \int_Y \left\{ \int_X f(x,y)\, d\mu(x) \right\} d\nu(y)?$$

An important example of this question arose at the end of the previous chapter. However, much preparation needs to be done in order to even ask the general question intelligently. For example, what sorts of functions $f(x,y)$ have the property that for *enough* fixed points x, the function $y \to f(x,y)$ is measurable on Y; and for *enough* fixed points y, the function $x \to f(x,y)$ is measurable on X? This question brings to light the fact that we will be dealing with *three* σ-algebras of sets here, one in X, another in Y, and a third in the product set $X \times Y$. Thus we begin with an investigation of *product* σ-algebras.

1. Product σ-algebras

Suppose that $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ are measurable spaces. A *measurable rectangle* is any set $R \in \mathcal{P}(X \times Y)$ having the form $R = A \times B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Definition 26. *$\mathcal{A} \times \mathcal{B}$ is the smallest σ-algebra on $X \times Y$ containing all measurable rectangles.*

An *elementary* set $E \in \mathcal{P}(X \times Y)$ is any finite pairwise disjoint union of measurable rectangles, i.e. $E = \dot{\bigcup}_{n=1}^N A_n \times B_n$ where $A_n \in \mathcal{A}$ and $B_n \in \mathcal{B}$. The collection of all elementary sets is denoted $\mathcal{E}$.

Definition 27. *A* monotone class $\mathcal{M}$ *on a set Z is a collection of sets in $\mathcal{P}(Z)$ that is closed under both monotone unions and monotone intersections, i.e.*

$$\begin{aligned} \bigcup_{n=1}^{\infty} E_n &\in \mathcal{M} \text{ if } E_n \in \mathcal{M} \text{ and } E_n \subset E_{n+1} \text{ for all } n \geq 1, \\ \bigcap_{n-1}^{\infty} E_n &\in \mathcal{M} \text{ if } E_n \in \mathcal{M} \text{ and } E_n \supset E_{n+1} \text{ for all } n \geq 1. \end{aligned}$$

Clearly every σ-algebra is also a monotone class. Since $\mathcal{A} \times \mathcal{B}$ contains the elementary sets $\mathcal{E}$, it follows thus $\mathcal{A} \times \mathcal{B}$ is a monotone class containing $\mathcal{E}$. It turns out that in order to define the notion of product measure *independent* of iteration, it is important that $\mathcal{A} \times \mathcal{B}$ is the *smallest* monotone class containing $\mathcal{E}$. Note that for *any* given collection of sets $\mathcal{F}$, the smallest monotone class containing $\mathcal{F}$ always exists - it is simply the intersection of all monotone classes containing $\mathcal{F}$.

THEOREM 59. *$\mathcal{A}\times\mathcal{B}$ is the smallest monotone class on $X\times Y$ containing the collection $\mathcal{E}$ of elementary sets.*

Proof: From the remarks made prior to the theorem we have

$$\mathcal{E}\subset\mathcal{M}\subset\mathcal{A}\times\mathcal{B}, \tag{1.1}$$

where $\mathcal{M}$ is the smallest monotone class containing $\mathcal{E}$. Now the intersection of two measurable rectangles is again a measurable rectangle, and the complement of a measurable rectangle is a union of three pairwise disjoint measurable rectangles, namely

$$\begin{aligned}(A_1\times B_1)\cap(A_2\times B_2) &= (A_1\cap A_2)\times(B_1\cap B_2),\\ (A_1\times B_1)^c &= (A_1^c\times B_1)\,\dot{\cup}\,(A_1\times B_1^c)\,\dot{\cup}\,(A_1^c\times B_1^c).\end{aligned}$$

From this we see that the collection $\mathcal{E}$ of elementary sets is closed under finite unions, intersections and differences, i.e.

$$P\cup Q,\ P\cap Q,\ P\setminus Q,\ Q\setminus P\in\mathcal{E} \text{ for all } P,Q\in\mathcal{E}. \tag{1.2}$$

Indeed, this is obvious for $P\cap Q$, and so then for $P\setminus Q=P\cap Q^c$, and finally then for $P\cup Q=(P\setminus Q)\,\dot{\cup}\,Q$.

Now for every $P\in\mathcal{P}(X\times Y)$ let

$$\mathcal{M}_P\equiv\{Q\in\mathcal{P}(X\times Y):P\setminus Q,\ Q\setminus P,\ P\cup Q\in\mathcal{M}\}.$$

It is clear that $\mathcal{M}_P$ is a monotone class for *every* $P\in\mathcal{P}(X\times Y)$, and moreover that

$$Q\in\mathcal{M}_P\Longleftrightarrow P\in\mathcal{M}_Q,\qquad \text{for all } P,Q\in\mathcal{P}(X\times Y). \tag{1.3}$$

We now claim that

$$P\setminus Q,\ Q\setminus P,\ P\cup Q\in\mathcal{M} \text{ for all } P,Q\in\mathcal{M}. \tag{1.4}$$

Indeed, suppose first that $P\in\mathcal{E}$. Then by (1.2) we have that $Q\in\mathcal{M}_P$ for all $Q\in\mathcal{E}$. Thus $\mathcal{E}\subset\mathcal{M}_P$ and hence also $\mathcal{M}\subset\mathcal{M}_P$ since $\mathcal{M}_P$ is a monotone class containing $\mathcal{E}$, and $\mathcal{M}$ is the smallest such. Now fix $Q\in\mathcal{M}$. We just proved that for $P\in\mathcal{E}$ we have $Q\in\mathcal{M}_P$, hence by (1.3) we also have $P\in\mathcal{M}_Q$. Thus $\mathcal{E}\subset\mathcal{M}_Q$, and hence also $\mathcal{M}\subset\mathcal{M}_Q$ since $\mathcal{M}_Q$ is a monotone class. This completes the proof of (1.4).

We next claim that $\mathcal{M}$ is a σ-algebra. Indeed, $\mathcal{M}$ is closed under complementation by (1.4) since if $P\in\mathcal{M}$, then $P^c=(X\times Y)\setminus P$ where both $X\times Y$ and P are in $\mathcal{M}$. Finally, $\mathcal{M}$ is closed under countable unions since if $\{P_n\}_{n=1}^\infty\subset\mathcal{M}$, then

$$\bigcup_{n=1}^\infty P_n=\bigcup_{n=1}^\infty\{P_1\cup P_2\cup\ldots\cup P_n\}\in\mathcal{M},$$

since the latter union is monotone and $P_1\cup P_2\cup\ldots\cup P_n\in\mathcal{M}$ for each $n\geq 1$ by (1.4).

In particular, we have proved that $\mathcal{M}$ is a σ-algebra containing the measurable rectangles. Since $\mathcal{A}\times\mathcal{B}$ is the smallest such we obtain $\mathcal{A}\times\mathcal{B}\subset\mathcal{M}$, which when combined with (1.1) gives $\mathcal{A}\times\mathcal{B}=\mathcal{M}$.

DEFINITION 28. *Given a function $f:X\times Y\to\mathbb{C}$ (or $[0,\infty]$), and a point $x\in X$, we define the* slice *function $f_x:Y\to\mathbb{C}$ (or $[0,\infty]$) by*

$$f_x(y)=f(x,y),\qquad y\in Y.$$

Similarly, for $y \in Y$*, we define the* slice *function* $f^y : X \to \mathbb{C}$ *(or* $[0,\infty]$*) by*

$$f^y(x) = f(x,y), \qquad x \in X.$$

Finally, for $E \in \mathcal{P}(X \times Y)$*, we define the* slices E_x *and* E^y *by*

$$\begin{aligned} E_x &= \{y \in Y : (x,y) \in E\}, \qquad x \in X, \\ E^y &= \{x \in X : (x,y) \in E\}, \qquad y \in Y. \end{aligned}$$

Note that

$$(\chi_E)_x = \chi_{E_x} \text{ and } (\chi_E)^y = \chi_{E^y}.$$

The *minimality* of the product σ-algebra $\mathcal{A}\times\mathcal{B}$ turns out to imply that measurability of $f(x,y)$ with respect to $\mathcal{A}\times\mathcal{B}$ is passed on to measurability of the slice functions f_x and f^y with respect to $\mathcal{B}$ and $\mathcal{A}$.

Theorem 60. *Let* f *be* $\mathcal{A}\times\mathcal{B}$*-measurable on* $X \times Y$*, and let* $E \in \mathcal{A}\times\mathcal{B}$*. Then*

1. *for every* $x \in X$*,* f_x *is* $\mathcal{B}$*-measurable on* Y*, and* $E_x \in \mathcal{B}$*;*
2. *for every* $y \in Y$*,* f^y *is* $\mathcal{A}$*-measurable on* X*, and* $E^y \in \mathcal{A}$*.*

Proof: Let V be an open set in $\mathbb{C}$ (or $[0,\infty]$) and let $G = f^{-1}(V)$. Then

$$(f_x)^{-1}(V) = \{y \in Y : (x,y) \in G\} = G_x.$$

Now let

$$\mathcal{C} = \{F \in \mathcal{A}\times\mathcal{B} : F_x \in \mathcal{B} \text{ for all } x \in X\}.$$

Since $\mathcal{B}$ and $\mathcal{A}\times\mathcal{B}$ are σ-algebras, it follows easily that $\mathcal{C}$ is a σ-algebra. Moreover, if $F = A \times B$ is a measurable rectangle, then $F_x = \left\{\begin{array}{ll} B & \text{if} \quad x \in A \\ \emptyset & \text{if} \quad x \in A^c \end{array}\right.$, and so $\mathcal{C}$ is a σ-algebra that contains all the measurable rectangles. We conclude that $\mathcal{C} = \mathcal{A}\times\mathcal{B}$. Since $G \in \mathcal{A}\times\mathcal{B} = \mathcal{C}$, we have $(f_x)^{-1}(V) = G_x \in \mathcal{B}$ for all $x \in X$, which shows that f_x is $\mathcal{B}$-measurable for all $x \in X$. In particular, $\chi_{E_x} = (\chi_E)_x$ is $\mathcal{B}$-measurable and $E_x \in \mathcal{B}$ for all $x \in X$. Similarly, f^y is $\mathcal{A}$-measurable and $E^y \in \mathcal{A}$ for all $y \in Y$.

2. Product measures

Let μ be a positive measure on $(X,\mathcal{A})$, and let ν be a positive measure on $(Y,\mathcal{B})$. In this section we consider the equality of the two natural candidates for defining a *product measure* $\mu \times \nu$ on $\mathcal{A}\times\mathcal{B}$, namely for $E \in \mathcal{A}\times\mathcal{B}$,

$$\int_X \left\{\int_Y (\chi_E)_x(y)\,d\nu(y)\right\} d\mu(x) \text{ and } \int_Y \left\{\int_X (\chi_E)^y(x)\,d\mu(x)\right\} d\nu(y). \tag{2.1}$$

We note that Theorem 60 shows that the functions $(\chi_E)_x$ and $(\chi_E)^y$ are measurable, and hence that the inner integrals $\int_Y (\chi_E)_x(y)\,d\nu(y)$ and $\int_X (\chi_E)^y(x)\,d\mu(x)$ in (2.1) exist for all x and y. But we don't yet know that the functions

$$x \to \int_Y (\chi_E)_x(y)\,d\nu(y) \text{ and } y \to \int_X (\chi_E)^y(x)\,d\mu(x)$$

are measurable, and so we can't yet make sense of the iterated integrals in (2.1).

However, even when we *can* make sense of both iterated integrals, they may *not* be equal! For example, if μ is Lebesgue measure on $(\mathbb{R},\mathcal{L}_1)$, and ν is counting measure on $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$, and $E = \{(x,x) : 0 \le x \le 1\}$ is a diagonal segment in $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, then

$$\int_X \left\{ \int_Y (\chi_E)_x (y)\, d\nu(y) \right\} d\mu(x) = \int_{[0,1]} \{1\}\, d\mu(x) = 1 \cdot 1 = 1,$$

and

$$\int_Y \left\{ \int_X (\chi_E)^y (x)\, d\mu(x) \right\} d\nu(y) = \int_{[0,1]} \{0\}\, d\nu(y) = 0 \cdot \infty = 0.$$

The following theorem resolves these difficulties when the measures μ and ν are both *σ-finite*, i.e. $X = \dot{\bigcup}_{i=1}^{\infty} X_i$ with $\mu(X_i) < \infty$ for all i, and $Y = \dot{\bigcup}_{j=1}^{\infty} Y_j$ with $\nu(Y_j) < \infty$ for all j. The proof will use the Monotone and Dominated Convergence Theorems in conjunction with Theorem 59 on monotone classes.

Theorem 61. *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces, and let $E \in \mathcal{A} \times \mathcal{B}$. Then*

$$\varphi(x) \equiv \int_Y (\chi_E)_x (y)\, d\nu(y) = \nu(E_x), \qquad x \in X,$$

is $\mathcal{A}$-measurable, and

$$\psi(y) \equiv \int_X (\chi_E)^y (x)\, d\mu(x) = \mu(E^y), \qquad y \in Y,$$

is $\mathcal{B}$-measurable. Moreover, we have the equality

$$\int_X \varphi(x)\, d\mu(x) = \int_Y \psi(y)\, d\nu(y).$$

Proof: If both measures μ and ν were *finite*, we could use the Monotone and Dominated Convergence Theorems to show that the class of all sets $E \in \mathcal{A} \times \mathcal{B}$ that satisfy the conclusions of the theorem, is a monotone class containing the elementary sets. We could then apply Theorem 59 to complete the proof of the theorem. Since the measures μ and ν are only σ-finite, we must be a bit more careful.

Let $\mathcal{C}$ be the class of all sets $E \in \mathcal{A} \times \mathcal{B}$ that satisfy the conclusions of the theorem. We claim that $\mathcal{C}$ has the following four properties:

(1) Every measurable rectangle $A \times B \in \mathcal{A} \times \mathcal{B}$ belongs to $\mathcal{C}$,
(2) If $\{E_n\}_{n=1}^{\infty}$ is a *nondecreasing* sequence of sets in $\mathcal{C}$, i.e. $E_n \subset E_{n+1}$ for all $n \ge 1$, then $E \equiv \bigcup_{n=1}^{\infty} E_n \in \mathcal{C}$,
(3) If $\{E_n\}_{n=1}^{\infty}$ is a *pairwise disjoint* sequence of sets in $\mathcal{C}$, i.e. $E_m \cap E_n = \emptyset$ for all $m, n \ge 1$, then $E \equiv \bigcup_{n=1}^{\infty} E_n \in \mathcal{C}$,
(4) Suppose that $A \times B$ is a measurable rectangle with $\mu(A) < \infty$ and $\nu(B) < \infty$. Then if $\{E_n\}_{n=1}^{\infty}$ is a *nonincreasing* sequence of sets in $\mathcal{C}$, i.e. $E_n \supset E_{n+1}$ for all $n \ge 1$, and if $A \times B \supset E_1$, then $E \equiv \bigcap_{n=1}^{\infty} E_n \in \mathcal{C}$.

With these four properties established for $\mathcal{C}$, it is easy to finish the proof of the theorem. Indeed, we simply define

$$\mathcal{M} \equiv \{E \in \mathcal{A} \times \mathcal{B} : E \cap (X_i \times Y_j) \in \mathcal{C} \text{ for all } i, j \ge 1\},$$

where X_i and Y_j are as in the definition of σ-finiteness of X and Y. Properties (2) and (4) show that $\mathcal{M}$ is a monotone class. Properties (1) and (3) show that the elementary sets $\mathcal{E}$ are contained in $\mathcal{M}$. Theorem 59 now shows that $\mathcal{M} = \mathcal{A} \times \mathcal{B}$, and the theorem is proved.

So it remains only to establish properties (1) through (4) for the class $\mathcal{C}$. If $E = A \times B$, then

$$(\chi_E)_x(y) = \chi_A(x)\,\chi_B(y) = (\chi_E)^y(x),$$

is $\mathcal{B}$-measurable for each x, and $\mathcal{A}$-measurable for each y, and so

$$\varphi(x) = \int_Y \chi_A(x)\,\chi_B(y)\,d\nu(y) = \nu(B)\,\chi_A(x) \text{ is measurable,}$$

$$\psi(y) = \int_X \chi_A(x)\,\chi_B(y)\,d\mu(x) = \mu(A)\,\chi_B(y) \text{ is measurable,}$$

$$\int_X \varphi(x)\,d\mu(x) = \mu(A)\,\nu(B) = \int_Y \psi(y)\,d\nu(y).$$

This establishes property (1).

To prove property (2), we let φ_n and ψ_n correspond to E_n in the same way that φ and ψ correspond to E above. We are assuming that φ_n and ψ_n satisfy the conclusions of the theorem, so they are both measurable and

$$\int_X \varphi_n(x)\,d\mu(x) = \int_Y \psi_n(y)\,d\nu(y), \qquad n \geq 1.$$

Since the sequence of sets E_n is nondecreasing, the sequence of functions φ_n is nondecreasing, and so is the sequence of functions ψ_n. The Monotone Convergence Theorem applied twice gives

$$\begin{aligned}\int_X \varphi(x)\,d\mu(x) &= \lim_{n\to\infty}\int_X \varphi_n(x)\,d\mu(x)\\ &= \lim_{n\to\infty}\int_Y \psi_n(y)\,d\nu(y) = \int_Y \psi(y)\,d\nu(y).\end{aligned}$$

This completes the proof of property (2).

Property (3) is obvious for finite pairwise disjoint unions, and the general case then follows using property (2).

Finally, the proof of property (4) is similar to that of property (2), except that we can use the Dominated Convergence Theorem instead of the Monontone Convergence Theorem because both $\mu(A)$ and $\nu(B)$ are finite.

We can now define the product measure $\mu \times \nu$ on $\mathcal{A} \times \mathcal{B}$ that is associated with μ and ν.

Definition 29. *If $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are σ-finite measure spaces, and if $E \in \mathcal{A} \times \mathcal{B}$, define*

(2.2)

$$\mu \times \nu(E) = \int_X \left\{\int_Y (\chi_E)_x(y)\,d\nu(y)\right\} d\mu(x) = \int_Y \left\{\int_X (\chi_E)^y(x)\,d\mu(x)\right\} d\nu(y),$$

where the equality of the two iterated integrals follows from Theorem 61.

Corollary 14 applied twice shows that $\mu\times\nu$ is a positive measure on $(X \times Y, \mathcal{A} \times \mathcal{B})$, and it is of course σ-finite. With the definition of product measure in hand, we are more than half way to proving the equality of iterated integrals in Fubini's theorem. Indeed, taking finite sums of scalars times indicator functions in (2.2) shows that

$$\int_X \left(\int_Y f\ d\nu\right) d\mu = \int_{X\times Y} f\ d(\mu\times\nu) = \int_Y \left(\int_X f\ d\mu\right) d\nu, \tag{2.3}$$

for all simple functions f. Five applications of the Monotone Convergence Theorem then show that (2.3) holds for nonnegative measurable f. The integrals on the far left and far right in (2.3) are called *iterated* integrals, and the integral in the middle is called a *double* integral. In the next section we give a precise and more general statement, along with a detailed proof. The cases where f is $[0,\infty]$-valued and $\mathbb{C}$-valued are treated separately.

3. Fubini's theorem

THEOREM 62. *Let $(X,\mathcal{A},\mu)$ and $(Y,\mathcal{B},\nu)$ be σ-finite measure spaces, and let f be $\mathcal{A}\times\mathcal{B}$-measurable on the product set $X\times Y$.*

(1) *If $0\leq f(x,y)\leq\infty$ for all $(x,y)\in X\times Y$, and if*

$$\begin{aligned}\varphi(x) &\equiv \int_Y f_x(y)\,d\nu(y), \qquad x\in X,\\ \psi(y) &\equiv \int_X f^y(x)\,d\mu(x), \qquad y\in Y,\end{aligned}$$

then φ is $\mathcal{A}$-measurable and ψ is $\mathcal{B}$-measurable and

$$\int_X \varphi\ d\mu = \int_{X\times Y} f\ d(\mu\times\nu) = \int_Y \psi\ d\nu.$$

(2) *If $f(x,y)\in\mathbb{C}$ for all $(x,y)\in X\times Y$, and*

$$\varphi^*(x)\equiv\int_Y |f|_x(y)\,d\nu(y), \qquad x\in X,$$

then

$$\int_{X\times Y}|f|\ d(\mu\times\nu)=\int_Y \varphi^*\ d\nu,$$

and so $f\in L^1(\mu\times\nu)$ if $\int_Y\varphi^*\ d\nu<\infty$.

(3) *If $f\in L^1(\mu\times\nu)$ then $f_x\in L^1(\nu)$ for μ-almost every $x\in X$, $f^y\in L^1(\mu)$ for ν-almost every $y\in Y$, the functions φ and ψ defined almost everywhere by*

$$\begin{aligned}\varphi(x) &\equiv \int_Y f_x(y)\,d\nu(y), \qquad \mu-a.e.\ x\in X,\\ \psi(y) &\equiv \int_X f^y(x)\,d\mu(x), \qquad \nu-a.e.\ y\in Y,\end{aligned}$$

are in $L^1(\mu)$ and $L^1(\nu)$ respectively, and

$$\int_X \varphi\ d\mu = \int_{X\times Y} f\ d(\mu\times\nu) = \int_Y \psi\ d\nu.$$

The first assertion (1) is often called Tonelli's Theorem, while the third assertion (3) is then referred to as Fubini's Theorem. The point of assertion (2) is that if *at least one* of the iterated integrals of $|f|$ is finite, then $f\in L^1(\mu\times\nu)$ and so (3) holds.

Proof: We first prove assertion (1). If $E\in\mathcal{A}\times\mathcal{B}$, then Theorem 61 shows that assertion (1) holds for $f=\chi_E$. By summing scalar multiples of such indicator functions, we see that (1) holds for all simple functions f. Now if f is $[0,\infty]$-valued, Proposition 7 shows that there is a nondecreasing sequence $\{s_n\}_{n=1}^\infty$ of nonnegative simple functions satisfying $0\leq s_n\leq s_{n+1}\leq f$ for all $n\geq 1$ and such that $\lim_{n\to\infty} s_n(x,y) = f(x,y)$ for every $(x,y) \in X \times Y$. Since assertion (1) holds for s_n, if we let φ_n and ψ_n correspond to s_n in the same way that φ and ψ correspond to f, then we have

$$\int_X \varphi_n \, d\mu = \int_{X\times Y} s_n \, d(\mu\times\nu) = \int_Y \psi_n \, d\nu, \qquad \text{for all } n \geq 1.$$

We now apply the Monotone Convergence Theorem five times. Two applications show that φ_n increases pointwise to φ, and that ψ_n increases pointwise to ψ. Three more applications show that

$$\begin{aligned} \lim_{n\to\infty} \int_X \varphi_n \, d\mu &= \int_X \varphi \, d\mu, \\ \lim_{n\to\infty} \int_{X\times Y} s_n \, d(\mu\times\nu) &= \int_{X\times Y} f \, d(\mu\times\nu), \\ \lim_{n\to\infty} \int_X \psi_n \, d\mu &= \int_X \psi \, d\mu, \end{aligned}$$

and this completes the proof of assertion (1).

Assertion (2) is an immediate consequence of applying assertion (1) to $|f|$.

Finally, assertion (3) is easily reduced to the case that f is real-valued. Assertion (1) then applies to both the positive f_+ and negative f_- parts of f to give

$$\int_X \varphi_\pm \, d\mu = \int_{X\times Y} f_\pm \, d(\mu\times\nu) = \int_Y \psi_\pm \, d\nu,$$

where $\varphi_\pm$ and $\psi_\pm$ correspond to $f_\pm$ in the same way that φ and ψ correspond to f. Now we *add* the two equations corresponding to $\pm$ to obtain that

$$\int_X |\varphi| \, d\mu = \int_{X\times Y} |f| \, d(\mu\times\nu) = \int_Y |\psi| \, d\nu.$$

Thus the functions $\varphi_\pm, f_\pm, \psi_\pm$ are finite almost everywhere, and all have finite integral. Thus we can take the *difference* of the two equations corresponding to $\pm$ to obtain

$$\int_X \varphi \, d\mu = \int_{X\times Y} f \, d(\mu\times\nu) = \int_Y \psi \, d\nu.$$

Note that the indeterminate expression $\infty - \infty$ will only arise on sets of measure zero in the differences taken above. This completes the proof of Fubini's theorem.

The next two examples show that assertion (3) of Fubini's theorem may fail if

- f is *not* integrable, even if all other hypotheses hold,
- f is *not* $\mathcal{A}\times\mathcal{B}$-measurable, even if all other hypotheses hold.

EXAMPLE 7. *Even if X and Y are* finite *measure spaces, f is $\mathcal{A}\times\mathcal{B}$-measurable, and both iterated integrals for f exist, it may happen that the iterated integrals are* not *equal, due to the fact that f fails to be integrable. For example, let $X = Y = [0,1)$, let $\mu = \nu$ be Lebesgue measure on $[0,1)$, and define f by*

$$f(x,y) = \sum_{n=1}^{\infty} \left\{ 2^{n+1} \chi_{\left[\frac{1}{2^{n+1}}, \frac{1}{2^n}\right)}(x) - 2^n \chi_{\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)}(x) \right\} 2^n \chi_{\left[\frac{1}{2^n}, \frac{1}{2^{n-1}}\right)}(y).$$

Then if $x \in \left[\frac{1}{2},1\right)$*, we have* $f_x(y) = -4\chi_{\left[\frac{1}{2},1\right)}(y)$ *and* $\int f_x(y)\,dy = -2$*, while if* $x \in \left[\frac{1}{2^{n+1}},\frac{1}{2^n}\right)$ *for some* $n \geq 1$*, then*

$$f_x(y) = 2^{n+1}2^n\chi_{\left[\frac{1}{2^n},\frac{1}{2^{n-1}}\right)}(y) - 2^{n+1}2^{n+1}\chi_{\left[\frac{1}{2^{n+1}},\frac{1}{2^n}\right)}(y),$$

and $\int f_x(y)\,dy = 0$*. Altogether we have*

$$\int\left(\int f_x(y)\,dy\right)dx = \int_{\left[\frac{1}{2},1\right)}(-2)\,dx + \sum_{n=1}^{\infty}\int_{\left[\frac{1}{2^{n+1}},\frac{1}{2^n}\right)} 0dx = -1.$$

On the other hand, we have $\int f^y(x)\,dx = 0$ *for all* $y \in [0,1)$ *and so*

$$\int\left(\int f^y(x)\,dx\right)dy = \int 0dy = 0.$$

EXAMPLE 8. *Even if* X *and* Y *are* finite *measure spaces, and both iterated integrals exist for a nonnegative bounded function* f*, it may happen that the iterated integrals are* not *equal, due to the fact that* f *fails to be* $\mathcal{A}\times\mathcal{B}$*-measurable. For example, let both* $(X,\mathcal{A},\mu)$ *and* $(Y,\mathcal{B},\nu)$ *be Lebesgue measure on* $[0,1]$*. Assume the axiom of choice, and in addition the* continuum hypothesis*, which asserts that the cardinality of the real numbers is the first uncountable cardinal. Then there is a one-to-one mapping*

$$\Gamma : [0,1] \to X \setminus \{\omega_1\},$$

where $(X,\prec)$ *is the well-ordered set whose last element is the first uncountable ordinal* ω_1*. See the fourth instance of a measure space in Example 4 near the beginning of Chapter 6. We note in passing that Cohen's famous theorem shows that the continuum hypothesis is* independent *of ZFC set theory, the Zermelo-Fraenkel axioms together with the axiom of choice. Now define*

$$E \equiv \left\{(x,y) \in [0,1]^2 : \Gamma(x) \prec \Gamma(y)\right\}.$$

Recall that there are at most countably many predecessors of α *for any* $\alpha \in X\setminus\{\omega_1\}$*. Thus for each* $x \in [0,1]$*, the slice* E_x *contains all but at most countably many of the points in* $[0,1]$*, and so is Borel measurable with measure* 1*. Also, for each* $y \in [0,1]$*, the slice* E^y *contains at most countably many of the points in* $[0,1]$*, and so is Borel measurable with measure* 0*. Thus the iterated integrals of* χ_E *both exist and we compute that*

$$\begin{aligned}\int_{[0,1]}\left(\int_{[0,1]}(\chi_E)_x(y)\ \ dy\right)dx &= \int_{[0,1]} 1dx = 1,\\ \int_{[0,1]}\left(\int_{[0,1]}(\chi_E)^y(x)\ \ dx\right)dy &= \int_{[0,1]} 0dy = 0.\end{aligned}$$

Bibliography

[1] R. G. Bartle and D. R. Sherbert, *Introduction to Real Analysis*, John Wiley and Sons, Inc. 3rd edition, 2000.

[2] C. B. Boyer, *A history of mathematics*, John Wiley & Sons, Inc., 1968.

[3] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, 3rd edition, 1976.

[4] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, 3rd edition, 1987.

[5] W. Rudin, *Functional Analysis*, International Series in Pure and Appl. Math., McGraw-Hill, 2nd edition, 1991.

[6] E. M. Stein and R. Shakarchi, *Complex Analysis, Princeton Lectures in Analysis* **II**, Princeton University Press, Princeton and Oxford, 2003.

[7] E. M. Stein and R. Shakarchi, *Real Analysis, Princeton Lectures in Analysis* **III**, Princeton University Press, Princeton and Oxford, 2003.

[8] S. Wagon, *The Banach-Tarski Paradox*, Cambridge University Press, 1985.