MATH3320 summary

Stanley Li

June 4, 2023

Contents

0 Notation, conventions and perspectives

The *characteristic function* χ_S or $\mathbb{1}_S$ of a set $S \subseteq X$ is the function

$$\chi_S = \mathbb{1}_S : X \to \mathbb{R}$$
$$x \mapsto \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{otherwise} \end{cases}$$

0.1 An analogy from real numbers to sets

As below (and as previously seen) we have two supplementary notions of a limit of a sequence of real numbers $(a_n)_{n=1}^{\infty}$, namely an "upper limit"

$$\limsup_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k = \inf_{n\in\mathbb{N}} \sup_{k\geq n} a_k$$

and a "lower limit"

$$\liminf_{n\to\infty} a_n = \lim_{n\to\infty} \sup_{k\geq n} a_k = \sup_{n\in\mathbb{N}} \inf_{k\geq n} a_k$$

The lattermost descriptions are independent of the notion of a limit and hence the notion of a metric, and only require the notion of a partial order[1] to state.

Another natural partial order is on the subsets of any set X (where in particular we should think of $X = \mathbb{R}^d$ as we will be almost exclusively doing analysis over Euclidean space), given by subset containment. This means that even though we have no notion of a "limit of sets" (due to the lack of suitable metric on a collection of sets), we can define a natural notion of an "upper limit" and "lower limit" to a sequence of sets, by first describing a suitable notion of supremum and infimum. Naturally, we can take

$$\inf_{i\in I} U_i = \bigcap_{i\in I} U_i$$

and

$$\sup_{i\in I} U_i = \bigcup_{i\in I} U_i$$

Then, following the above descriptions of $\limsup$ and $\liminf$, for a sequence $(U_n)_{n=1}^{\infty}$ we naturally arrive at

$$\limsup_{n\to\infty} U_n = \inf_{n\in\mathbb{N}} \sup_{k\geq n} U_n = \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} U_n$$

and

$$\liminf_{n\to\infty} U_n = \sup_{n\in\mathbb{N}} \inf_{k\geq n} U_n = \bigcup_{n\in\mathbb{N}} \bigcap_{k\geq n} U_n$$

and it is probably unsurprising from these notions that $\liminf_{n\to\infty} U_n \subseteq \limsup_{n\to\infty} U_n$. These are the notions of $\limsup$ and $\liminf$ we take in the remainder of these notes. For an increasing sequence of sets $(U_n)_{n=1}^{\infty}$ we can treat its limsup as its actual limit, and the same for a decreasing sequence of sets and its liminf[2].

[1] A *partial order* $\preceq$ on a set X is a relation satisfying reflexivity ($x \preceq x$), antisymmetry ($x \preceq y$ and $y \preceq x$ if and only if $x = y$) and transitivity ($x \preceq y$ and $y \preceq z$ necessarily yields $x \preceq z$). This can be thought of as a sensible notion of "ordering".

[2] And in theory if these two limiting notions are equal, it could be possible to define the "limit" of a sequence of sets.

0.2 An overarching idea for Lebesgue integration

The Riemann integration approach is to attempt to directly measure the area beneath the graph, using overestimates and underestimates and attempting to reach a common ground between the two.

In Lebesgue integration we also look at approximations to functions, but we take a different approach. Rather than directly taking a *measure* and attempting to compute upper and lower areas, we instead start from the level of sets to be measured. Measuring the size of sets in this case does not attempt to assign every set a "volume", but rather only the ones which are well-approximated by open sets (equivalently closed sets, and notably finite unions of cubes).

Once we have established which sets can be assigned a reasonable volume, we move to considering which functions can be easily assigned a reasonable notion of an integral, and we find that such functions are "very close" to being continuous. We then use these functions as a building block to define the integral for any arbitrary function, starting from the simplest functions (where we get finite sums), then the previous functions with a reasonable notion of an integral, and then arbitrary functions.

1 Analysis review

Definition 1 (liminf, limsup)**.** *Let $(a_n)_{n=1}^{\infty} \in \mathbb{R}^{\infty}$ be a sequence.*

- *If (a_n) is bounded above, its* limsup *is*

$$\limsup_{n\to\infty} a_n := \lim_{n\to\infty} \sup_{k\geq n} a_k = \inf_{n\geq 1} \left(\sup_{k\geq n} a_k \right)$$

- *If (a_n) is bounded below, its* liminf *is*

$$\liminf_{n\to\infty} a_n := \lim_{n\to\infty} \inf_{k\geq n} a_k = \sup_{n\geq 1} \left(\inf_{k\geq n} a_k \right)$$

Notice that $\lim_{n\to\infty} a_n$ exists iff

$$\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$$

Definition 2 (Partitions)**.** *A* partition *of $[a,b] \subseteq \mathbb{R}$ is a set $\{x_0, \ldots, x_n\} \subseteq [a,b]$ with $a = x_0 < \ldots < x_n = b$.*

Definition 3 (Upper and lower areas; integrals)**.** *Let $f : [a,b] \to \mathbb{R}$ be bounded, and $P = \{x_0, \ldots, x_n\}$ be a partition of $[a,b]$.*

1.
- *The* lower area of f with respect to P *is*

$$L(P,f) := \sum_{i=0}^{n-1} \inf_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1} - x_i)$$

- *The* upper area of f with respect to P *is*

$$U(P,f) := \sum_{i=0}^{n-1} \sup_{x\in[x_i,x_{i+1}]} f(x)(x_{i+1} - x_i)$$

2.
   - *The* lower Riemann integral *of f is*
$$\underline{\int_{[a,b]}^{R}} f := \sup_P L(P,f)$$
   - *The* upper Riemann integral *of f is*
$$\overline{\int_{[a,b]}^{R}} f := \inf_P U(P,f)$$

3. *If $\underline{\int_{[a,b]}^{R}} f = \overline{\int_{[a,b]}^{R}} f$, we say that f is* Riemann integrable*, and define*
$$\int_{[a,b]}^{R} f := \underline{\int_{[a,b]}^{R}} f = \overline{\int_{[a,b]}^{R}} f$$

Definition 4 (Rectangles, volume)**.** *A rectangle $R \subseteq \mathbb{R}^d$ is of the form*
$$R := \prod_{i=1}^{d} [a_i, b_i] = [a_1, b_1] \times \ldots \times [a_d, r_d]$$
with volume
$$|R| := \prod_{i=1}^{d} (b_i - a_i)$$

A *cube* is rectangle with all sidelengths equal.

2 Measure theory

2.1 Simple functions

Rather than partitioning the domain as with Riemann integration, we instead opt to partition the range. The usual approximations of a bounded $f : [a,b] \to \mathbb{R}$ are by the functions
$$\varphi_{L,P} := \sum_{i=0}^{n-1} \inf_{x \in [x_i, x_{i+1}]} \mathbb{1}_{[x_i, x_{i+1}]}$$
$$\varphi_{U,P} := \sum_{i=0}^{n-1} \sup_{x \in [x_i, x_{i+1}]} \mathbb{1}_{[x_i, x_{i+1}]}$$
We instead opt for the following setup: Let $f([a,b]) \subseteq [c,d]$ and $Q = \{y_0, \ldots, y_n\}$ be a partition of $[c,d]$ (that is, $c = y_0 < \ldots < y_n = d$). Then we define
$$\varphi_Q := \sum_{i=0}^{n} y_i \mathbb{1}_{f^{-1}[y_i, y_{i+1}]}$$
To define the Riemann integral from this perspective, we want a notion of the "measure" of a set of the form $f^{-1}[y_i, y_{i+1}]$.

2.2 Elementary sets and Jordan measure

We take $A\dot{\sqcup}B$ to mean that A and B are "almost disjoint", in the sense that $\text{int}(A)$ and $\text{int}(B)$ are disjoint.

Definition 5 (Elementary sets; Jordan measure)**.** *Let $C \subseteq \mathbb{R}^d$. Then C is an* elementary set *if there are rectangles $R_1, \ldots, R_n$ with*

$$C = R_1\dot{\sqcup}\ldots\dot{\sqcup}R_n$$

1. *If C is elementary as above, its* Jordan measure *(or volume) is*

$$|C| := m^J(C) := \sum_{i=1}^{n} |R_i|$$

2. *For $S \subseteq \mathbb{R}^d$,*
   - *the* inner Jordan measure $m_*^J(S)$ *is*

$$m_*^J(S) := \sup_{C\subseteq S \text{ elementary}} |C|$$

   - *the* outer Jordan measure $m_J^*(S)$ *is*

$$m_J^*(S) := \inf_{C\supseteq S \text{ elementary}} |C|$$

3. *If $m_*^J(S) = m_J^*(S)$, we say S is* Jordan measurable, *and define its* Jordan measure $m^J(S)$ *as $m^J(S) := m_*^J(S) = m_J^*(S)$.*

By rewriting the definitions, we have

Theorem 1 (Classification of Jordan measurable sets)**.** *Let $B \subseteq \mathbb{R}^d$ be bounded. Then B is Jordan measurable if and only if $\mathbb{1}_B$ is Riemann integrable.*

We want a notion of integrability for which we have some notion of commutativity with integrals and limits. For $\mathbb{Q} \cap [0, 1] = \{q_1, q_2, \ldots\}$, this means

$$\lim_{n\to\infty} \mathbb{1}_{\{q_1,\ldots,q_n\}} = \mathbb{1}_{\mathbb{Q}\cap[0,1]}$$

and also in the context of integrals.

Theorem 2 (Coverings of open sets)**.** *Every open set $O \subseteq \mathbb{R}^d$ is the almost disjoint union of countably many cubes in $\mathbb{R}^d$.*

The key idea is to restrict to consider O and its relative position to the lattices $\mathbb{Z}^d$, $\left(\frac{1}{2}\mathbb{Z}\right)^d, \ldots$. In 1-dimension, we can look at $O \cap \mathbb{Q} = \{q_1, q_2, \ldots\}$ and take the largest intervals O_i contained in O containing q_i.

2.3 Exterior measure

Definition 6 (Exterior measure)**.** *Let $E \subseteq \mathbb{R}^d$. Its* exterior measure *is*

$$m_*(E) = \inf \left\{ \sum_{i=1}^{k} |Q_i| \;\middle|\; k \in \mathbb{N} \cup \{\infty\}, E \subseteq \bigcup_{i=1}^{k} Q_i \right\}$$

Here m_* is a map $m_* : \mathcal{P}(\mathbb{R}^d) \to [0, \infty]$. The exterior measure has the following properties:

1. If E is countable, then $m_*(E) = 0$;
2. if $A \subseteq B$, $m_*(A) \leq m_*(B)$;
3. $m_*(\mathbb{R}^d) = \infty$; and
4. if Q is a (closed or open) cube, then $m_*(Q) = |Q|$.

We further have

1. **[Volume of rectangles]** If R is a rectangle, then $m_*(R) = |R|$
2. **[Countable additivity]** If $E \subseteq \bigcup_{j=1}^{\infty} E_j$, then

$$m_*(E) \leq \sum_{j=1}^{\infty} m_*(E_j)$$

3. **[Relation to open sets]** Let $\mathcal{O}_E = \{O \supseteq E \mid O \text{ is open}\}$. Then

$$m_*(E) = \inf_{O \in \mathcal{O}_E} m_*(O)$$

4. **[Exact additivity]** If $E = E_1 \sqcup E_2$ with $d(E_1, E_2) > 0$, then $m_*(E) = m_*(E_1) + m_*(E_2)$.
5. **[Almost disjoint unions of cubes]** If $E = \dot{\bigsqcup}_{j=1}^{\infty} Q_j$, then $m_*(E) = \sum_{j=1}^{\infty} |Q_j|$.

The first statement can intuitively be rephrased as saying that the surface area of a rectangle is $(d-1)$-dimensional, while its volume is d-dimensional, and so the ratio between these two quantities tends towards zero.

Since every open set is an almost disjoint union of cubes, we have

Corollary 1. *If $O_1, O_2, \ldots$ are disjoint open sets, then*

$$m_*\left(\bigsqcup_{j=1}^{\infty} O_j\right) = \sum_{j=1}^{\infty} m_*(O_j)$$

The key idea in the proof of these statements is to use the ε-criterion to show that a value is indeed the infimum of a set (namely to get the harder inequality). Namely, for each $\varepsilon > 0$, note that there is some element $s \in S$ with $s \leq \inf(S) + \varepsilon$, and try to relate your claimed value i_s with s (usually by showing $i_s \leq s$). It often also helps to write $\varepsilon = \varepsilon/2 + \varepsilon/4 + \ldots + \varepsilon/2^k + \ldots$.

2.4 Measurable sets

The ideal properties for a measure defined on all sets to satisfy are incompatible. We describe Lebesgue measurable sets, which are intuitively those which can be well-approximated (in terms of volume) by open sets.

Definition 7 (Lebesgue measurable)**.** *We say $E \subseteq \mathbb{R}^d$ is* Lebesgue measurable *if for $\varepsilon > 0$, we can find $O \supseteq E$ with*

$$m_*(O - E) < \varepsilon$$

In this case, its Lebesgue measure *is $m(E) := m_*(E)$.*

The set of measurable sets satisfies the following properties.

Theorem 3. *The following sets are measurable.*

1. *Any open $O \subseteq \mathbb{R}^d$.*
2. *Any set $E \subseteq \mathbb{R}^d$ of exterior measure zero.*
3. *A countable union of measurable sets.*
4. *Any closed $C \subseteq \mathbb{R}^d$.*
5. *The complement of any measurable set.*
6. *A countable intersection of measurable sets.*

By *(1.)*, *(3.)* and *(5.)*, Lebesgue measurable sets are thus a σ-algebra, as defined below.

Definition 8 (σ-algebra)**.** *A* σ-algebra *on $\mathbb{R}^d$ is a collection of subsets Σ closed under complement and countable union (and hence countable intersection) with $\emptyset \in \Sigma$.*

We note also that for any measurable set E, taking $\varepsilon = 1/n$ and O_n to be the associated open set, we find that for $O = \bigcap_{n=1}^{\infty} O_n$, $m_*(O - E) = 0$, and so the countable intersection O of open sets differs from E by a set of measure zero, i.e. $O = E \sqcup (O - E)$. We refer to such a set (i.e. a countable intersection of open sets) as a *G_δ-set.*

Similarly, doing the same for E^c, we find open sets $O_n \supseteq E^c$ with $m_*(O_n - E^c) < \frac{1}{n}$, i.e. $m_*(E - O_n^c) < \frac{1}{n}$, and then for $C = \bigcap_{n=1}^{\infty} O_n^c$ (which is a countable intersection of closed sets), we see that $m_*(E - C) = 0$, i.e. E differs from a countable union of closed sets by a set of measure zero ($E = C \sqcup (E - C)$). A countable union of closed sets is referred to as a *F_σ-set.*

Hence all Lebesgue measurable sets can be "approximated" by G_δ and F_σ sets. We denote the set of Lebesgue measureable sets by $\mathcal{M}(\mathbb{R}^d)$

Other examples of σ-algebras are the trivial σ-algebra $\{\emptyset, \mathbb{R}^d\}$, the discrete σ-algebra $2^{\mathbb{R}^d} := \mathcal{P}(\mathbb{R}^d)$, and the Borel σ-algebra[3] $\mathcal{B}(\mathbb{R}^d)$, generated by the open sets in $\mathbb{R}^d$.

We have strict inclusions

$$\mathcal{B}(\mathbb{R}^d) \subsetneq \mathcal{M}(\mathbb{R}^d) \subsetneq \mathcal{P}(\mathbb{R}^d)$$

Indeed, for the second inclusion, we can use the axiom of choice to define a map $[0,1]/\sim \to [0,1]$ where $x \sim y \iff x - y \in \mathbb{Q}$, and this defines a non-measurable set.

For the first inclusion, we can take two Cantor sets – one with positive measure and one with measure 0, and describe a homeomorphism between them. Taking a non-measurable set contained in the Cantor set of positive measure (justified by a later result), its image is necessarily a measurable, non-Borel set.

The Lebesgue measure satisfies some translation invariance properties. Indeed, for any measurable $E \subseteq \mathbb{R}^d$, we have

1. [**Additive translation invariance**] For any $h \in \mathbb{R}^d$, $E + h$ is measurable with

$$m(E) = m(E + h)$$

2. [**Multiplicative translation invariance**] For any $\delta = (\delta_1, \ldots, \delta_d) \in (\mathbb{R}^+)^d$, $m(\delta E)$ (defined by pointwise multiplication) is measurable with

$$m(\delta E) = \left(\prod_{i=1}^{d} \delta_i \right) m(E)$$

[3]The Borel σ-algebra $\mathcal{B}(X)$ is a construction which works on any topological space X.

Unlike the exterior measure, due to our restriction to measurable sets, we have the following additivity with the Lebesgue measure.

Remark 1. *The Lebesgue measure is a measure μ on $\mathbb{R}^d$ with the properties that*

1. *(Open sets are "large") $\mu(O) > 0$ for any open O;*
2. *(Compact sets are "small") $\mu(K) < \infty$ for any compact K; and*
3. *(Additive translation invariance) $\mu(E + h) = \mu(E)$ for any measurable E.*

Such a measure is called a Haar measure, *and as $\mathbb{R}^d$ is a "nice" space (i.e. a locally compact additive topological group), it is actually the unique measure (up to a scalar) on $\mathbb{R}^d$ with these properties.*

Theorem 4 (Countable additivity)**.** *Let $E_j \subseteq \mathbb{R}^d$ be pairwise disjoint measurable sets (for $j \in \mathbb{N}$). Then*

$$m\left(\bigsqcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} m(E_j)$$

By the construction of a non-measurable set, we also see that the following holds.

Proposition 1. *Any set of positive measure contains a non-measurable subset.*

We have the following notion of "convergence" of subsets of $\mathbb{R}^d$.

Definition 9 (Convergence of sets)**.**
1. *Let $\{E_i\}_{i=1}^{\infty}$ be an increasing sequence of sets. We say that E_j* converges to *E and write $E_i \nearrow E$ if $E = \bigcup_{i=1}^{\infty} E_i$.*
2. *Let $\{E_i\}_{i=1}^{\infty}$ be a decreasing sequence of sets. We say that E_j* converges to *E and write $E_i \searrow E$ if $E = \bigcap_{i=1}^{\infty} E_i$.*

We can view these as suprema and infima of the associated sequences of sets. The Lebesgue measure has the following compatibility with this notion of set limit.

Lemma 1.
1. *Let $\{E_i\}_{i=1}^{\infty}$ be an increasing sequence of sets converging to E. Then $m(E) = \lim_{n\to\infty} m(E_n)$.*
2. *Let $\{E_i\}_{i=1}^{\infty}$ be a decreasing sequence of sets converging to E, and suppose that $E_k < \infty$ for some $k \in \mathbb{N}$. Then $m(E) = \lim_{n\to\infty}(E_n)$.*

We thus have the following characterisations of measurable sets.

Theorem 5. *Let $E \subseteq \mathbb{R}^d$ be measurable. Then for every $\varepsilon > 0$,*

1. *(Definition) there is an open set $O \supseteq E$ with $m(O - E) < \varepsilon$.*
2. *(Equivalent definition) There is a closed subset $C \subseteq E$ with $m(E - C) < \varepsilon$.*
3. *If $m(E) < \infty$, then there is an compact set $K \subseteq E$ with $m(E - K) < \varepsilon$.*
4. *If $m(E) < \infty$, then there is a finite almost-disjoint union of cubes[4] F with $m(E \triangle F) < \varepsilon$.*

2.5 Measurable functions

Having described measurable sets, we move to describing measurable functions, en route to defining Lebesgue integration.

[4] where $\triangle$ denotes symmetric difference.

Definition 10 (Measurable function). *Let $E \subseteq \mathbb{R}^d$ be measurable and $f : E \to \mathbb{R}$ be a function. We say that f is* measurable *if the preimages $f^{-1}[-\infty, a)$ (denoted $\{f < a\}$) are measurable for all $a \in \mathbb{R}$.*

By the properties of measurable sets, it is equivalent to say that any of $\{f \leq a\} := f^{-1}[-\infty, a]$, $\{f > a\} := f^{-1}(a, \infty]$, $\{f \geq a\} := f^{-1}[a, \infty]$, $f^{-1}(a, b)$, $f^{-1}[a, b]$ are measurable for all a, b.

For examples of measurable functions, we see that any continuous f is measurable as $f^{-1}[-\infty, a)$ is open in each case, and that for a measurable set E, its associated characteristic function χ_E is also measurable.

Measurable functions have the following compatibility with arbitrary functions.

Theorem 6. *Let $(f_n)_{n=1}^{\infty}$ be a sequence of measurable functions. Then*

$$\liminf_{n \to \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{k \geq n} f_k \qquad \text{and} \qquad \limsup_{n \to \infty} f_n = \inf_{n \in \mathbb{N}} \sup_{k \geq n} f_k$$

are measurable.

We then have the following corollary.

Corollary 2. *If $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions with $f_n \to f$, then f is measurable.*

We also have the following "closure" properties of measurable functions under (pointwise) addition and multiplication.

Theorem 7. *Let f and g be measurable functions. Then*

1. *f^k is measurable for any $k \in \mathbb{N}$.*
2. *$f + g$ and fg are measurable.*

We evidently have (2) $\implies$ (1), but the statement of (1) is used to prove (2), by using the fact that $fg = \frac{1}{4}((f+g)^2 - (f-g)^2)$.

Remark 2. *Algebraically, **Theorem 7** shows that the set of measurable functions on $\mathbb{R}^d$ is a subring of the ring of functions $\mathbb{R}^d \to \mathbb{R}$ with pointwise operations, which in turn contains the subring of continuous functions (i.e. those where preimages of open sets are open sets).*

We have the following "equivalence relation" between almost-equal functions.

Definition 11 (Equality almost everywhere). *Let $E \subseteq \mathbb{R}^d$ be measurable and $f, g : E \to \mathbb{R}$. We say that f and g are* equal almost everywhere *if*

$$m(\{x \in E \mid f(x) \neq g(x)\}) = 0$$

that is, they differ by a measure zero set.

From the perspective of measure theory, the statement that f and g are equal almost everywhere should be viewed as being "good enough" for all intensive purposes. The following proposition is one such way to make this notion explicit.

Proposition 2. *Let f and g be functions which are equal almost everywhere. Then if f is measurable, so is g.*

We will define integrals later in several stages, beginning with relatively basic functions and then extending these definitions to fit larger (and possibly more cathartic) classes of functions. For this, we have the following notion of approximating a measurable function f by an increasing sequence of step functions.

Proposition 3. *Let f be a non-negative measurable function. Then there exists a sequence of simple functions $\varphi_k \nearrow f$ pointwise.*

The previous result applies only to non-negative functions, and we have a result by writing $f = f^+ - f^-$ for non-negative functions $f^\pm$.

Theorem 8. *For each measurable f, there is a sequence of simple functions φ_k with*

$$|\varphi_k(x)| \leq |\varphi_{k+1}(x)|$$

and $\varphi_k \to f$ pointwise.

By instead replacing simple functions we have a slightly weaker approximation (i.e. for which this convergence happens *almost everywhere*).

Theorem 9. *For each measurable f, there is a sequence of step functions φ_k with $\lim_{k\to\infty} \varphi_k = f$ almost everywhere.*

2.6 Littlewood's three principles

We have the following "principles" for approximating functions by rectangles and their associated step functions.

1. Every measurable $E \subseteq \mathbb{R}^d$ is *nearly* a finite union of cubes;
2. Every measurable function f of finite-value is *nearly* continuous; and
3. Every convergent sequence of measurable functions is *nearly* uniformly convergent.

The first principle corresponds to the last statement in **Theorem 5**, and the second and third principles are statements known as Lusin's and Egorov's theorems respectively.

Theorem 10 (Egorov)**.** *Let $(f_k : E \to \mathbb{R})_{k=1}^\infty$ each be measurable, with $f_n \to f$ and $m(E) < \infty$. Then for every $\varepsilon > 0$, there is a closed set $A_\varepsilon \subseteq E$ with $m(E \setminus A_\varepsilon) < \varepsilon$ and $f_k \to f$ uniformly on A_ε.*

Egorov's theorem is extremely powerful, as the weak notion of pointwise convergence is near useless in comparison to uniform convergence, which preserves many important properties (such as integrals and (uniform) continuity). Having established Egorov's theorem, we are then able to prove

Theorem 11 (Lusin)**.** *Let $f : E \to \mathbb{R}$ be measurable and $m(E) < \infty$. Then for every $\varepsilon > 0$ there is a closed set A_ε with $m(E \setminus A_\varepsilon) < \varepsilon$ and $f|_{A_\varepsilon} : A_\varepsilon \to \mathbb{R}$ a continuous function.*

3 Integration theory

Equipped with the notions of measurable sets and functions move to properly defining the notions of integration in a more generalised manner (but only for measurable functions, for which this notion makes sense). This is done in stages, beginning with "basic" functions where integrals are just finite sums and building up to higher generality. In this section all functions are $\mathbb{R}^d \to \mathbb{R}$ unless otherwise stated. From the perspective of integration, every measurable function $f : E \to \mathbb{R}$ can be made into a function $\mathbb{R}^d \to \mathbb{R}$ by extending it to $\mathbb{R}^d$ by zero. This is denoted $f\chi_E$, and we treat $\int_E f = \int_{\mathbb{R}^d} f\chi_E$ (when these are defined).

3.1 Lebesgue integration

3.1.1 Stage 1 – Simple functions

Let φ be a simple function, i.e. of the form $\varphi = \sum_{k=1}^n a_k\chi_{E_k}$ where E_k are measurable sets. Its Lebesgue integral is then

$$\int_{\mathbb{R}^d} \varphi := \int_{\mathbb{R}^d} \varphi(x)dx := \sum_{k=1}^n a_k m(E_k)$$

This is well-defined (independent of the choice of measurable sets and coefficients a_k), and every such simple function has a unique form $\varphi = \sum_{k=1}^n a_k\chi_{E_k}$ where a_k are distinct and E_k are pairwise disjoint.

3.1.2 Stage 2 – Bounded measurable functions

When we say "bounded, measurable" in this sense we mean a function whose domain and range are both bounded in some sense (not just its range). We have the following lemma which says that we can approximate the integrals of bounded measurable functions by those of simple functions.

Lemma 2. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function with $m(\mathrm{supp}(f)) < \infty$, and let $\varphi_k(x) \to f(x)$ everywhere in $\mathbb{R}^d$. Then*

1. *The limit*
$$\lim_{k\to\infty} \int_E \varphi_k$$
exists; and
2. *If $\psi_k(x) \to f(x)$ everywhere in E, then*
$$\lim_{k\to\infty} \left(\int_E (\varphi_k - \psi_k) \right) = 0$$
and in particular $\lim_{k\to\infty} \int_E \varphi_k = \lim_{k\to\infty} \int_E \psi_k$.

We then define

$$\int_{\mathbb{R}^d} f := \lim_{n\to\infty} \int_{\mathbb{R}^d} \varphi_k$$

for any sequence of simple functions $\varphi_k \to f$ pointwise, and the above lemma shows that this is well-defined. We also have the first integral convergence theorem, applicable to bounded functions with "bounded" domains.

Theorem 12 (Bounded convergence theorem)**.** *Let (f_n) be a sequence of bounded, measurable functions supported on a measurable set E of finite measure. Suppose also that $f_n \to f$ pointwise almost everywhere. Then f is measurable, bounded, and $m(\mathrm{supp}(f) \triangle E) = 0$, and*

$$\int |f_n - f| \to 0$$

and in particular $\int f_n \to \int f$.

3.1.3 Stage 3 – Positive measurable functions

We define integrals in this case by taking the "largest" of the underestimates by bounded measurable functions. Namely, we set

$$\int_{\mathbb{R}^d} f := \sup \left\{ \int_{\mathbb{R}^d} g \;\middle|\; 0 \leq g \leq f, g \text{ bounded and measurable} \right\}$$

and we say that f (when positive) is integrable when $\int_{\mathbb{R}^d} f$ We have the following properties of integrals of positive, measurable functions.

Lemma 3. *Let $f, g \geq 0$ be measurable, E be measurable and $\lambda, \mu \in \mathbb{R}$.*

1. *(Linearity)*
$$\int_E (\lambda f + \mu g) = \lambda \int_E f + \mu \int_E g$$
2. *(Additivity) If E_1, E_2 are disjoint and measurable, then*
$$\int_{E_1 \sqcup E_2} f = \int_{E_1} f + \int_{E_2} f$$
3. *(Monotonicity) If $f \leq g$ on E, then*
$$\int_E f \leq \int_E g$$
Further, if g is integrable, then so if f.
4. *(Integrability) If f is integrable, then $m(f^{-1}(\infty)) = 0$.*
5. *(Zero) If $\int_{\mathbb{R}^d} f = 0$, then $f = 0$ almost everywhere.*
6. *(Triangle inequality)*
$$\left| \int_E f \right| \leq \int_E |f|$$

3.1.4 Stage 4 – General measurable functions

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Writing $f = f^+ - f^-$ where $f^\pm = \max(\pm f, 0)$, we set its integral to be

$$\int_{\mathbb{R}^d} f := \int_{\mathbb{R}^d} f^+ - \int_{\mathbb{R}^d} f^-$$

To check that this is well-defined, we have that if $f = f_1 - f_2 = g_1 - g_2$ for positive f_i, g_i, then $\int f_1 + \int g_2 = \int g_1 + \int f_2$, so the value of the above definition is independent of the choice of decomposition into positive functions. Such a function is integrable if $\int_{\mathbb{R}^d} |f| < \infty$.

Properties (1) – (4) of Lemma 3 also hold for general measurable functions. In terms of integrals, all convergence will be pointwise almost everywhere.

3.2 Convergence of integrals

The first result we have relates the integrals of f_n and f for $f_n \to f$.

Proposition 4 (Fatou Lemma). *Let (f_n) be a sequence of non-negative measurable functions with $f_n(x) \to f(x)$ almost everywhere. Then*

$$\int f \leq \liminf_{n \to \infty} \int f_n$$

This can be thought of as an artifact of the fact that integrals are defined in terms of underestimates.

Corollary 3. *Let (f_n) be a sequence of non-negative measurable functions.*

1. *If $f_n(x) \to f(x)$ and $f_n(x) \leq f(x)$ for almost every x, then*
$$\int f = \lim_{n\to\infty} \int f_n$$
2. *If $a_k \geq 0$ are integrable functions, then*
$$\int \sum_{k=1}^{\infty} a_k = \sum_{k=1}^{\infty} \int a_k$$
which may be infinite. Further, if $\sum_{k=1}^{\infty} \int a_k$ is finite, then $\sum_{k=1}^{\infty} a_k$ converges for almost every x.
3. *(Monotone convergence theorem) If (f_n) are non-negative and $f_n(x) \nearrow f(x)$ pointwise almost everywhere, then*
$$\int f_n \to \int f$$
4. *(Borel-Cantelli Lemma) Let $\{E_i\}_{i=1}^{\infty}$ be a sequence of measurable sets with $\sum_{i=1}^{\infty} m(E_i) < \infty$. Then*
$$m\left(\limsup_{n\to\infty} E_n\right) = m\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} E_n\right) = 0$$
5. *For any $\delta > 0$, the function $f(x) = 1/|x|^{d+\delta}$ on $\mathbb{R}^d$ is integrable on $|x| \geq \varepsilon$ for any $\varepsilon > 0$.*

We have the following (rather strong) convergence theorem on integrals. It implies the bounded convergence theorem, but not the monotone convergence theorem, as it only deals with finite integrals.

Theorem 13 (Dominated convergence theorem)**.** *Suppose that (f_n) are integrable and f is such that $f_n(x) \to f(x)$ almost everywhere. If there is an integrable function g such that $|f_n(x)| \leq g(x)$ for almost every x, then*
$$\int |f_n - f| \to 0$$
and in particular
$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n$$

The proof of this theorem requires the following results.

Lemma 4. *Let g be an integrable function, and $\varepsilon > 0$.*

1. *There is $N > 0$ such that*
$$\int_{B_N(0)^c} |g| < \varepsilon$$
2. *(Absolute continuity) There is $\delta > 0$ so that*
$$\int_E |g| < \varepsilon$$
whenever $m(E) < \delta$.

We have the following relationship between Riemann and Lebesgue integration.

Proposition 5. *Let f be a Riemann integrable function. Then f is Lebesgue integrable, and its Riemann and Lebesgue integrals are equal.*

In this sense Lebesgue integration is more general than and compatible with Riemann integration.

3.3 $L^1(\mathbb{R}^d)$ and its properties

Definition 12 ($L^1(\mathbb{R}^d)$)**.** *The* space $L^1(\mathbb{R}^d)$ of integrable functions on $\mathbb{R}^d$ *is*

$$L^1(\mathbb{R}^d) = \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; \int_{\mathbb{R}^d} |f| < \infty \right\} \Big/ \sim$$

where $f \sim g$ *iff* $f = g$ *almost everywhere.*

This set is naturally a normed vector space over $\mathbb{R}$ with the L^1 *norm*

$$\|f\|_{L^1} := \int_{\mathbb{R}^d} |f|$$

In fact, it is complete (i.e. a Banach space):

Theorem 14 (Riesz-Fischer)**.** $L^1(\mathbb{R}^d)$ *is complete with the* L^1*-norm.*

In general, L^1 and pointwise convergence have no relation, but we have the following relation.

Corollary 4. *Every* L^1*-convergent sequence contains a pointwise convergent subsequence.*

It is usually convenient to view a metric space through the lens of a "much smaller" dense subset (such as $\mathbb{Q} \subseteq \mathbb{R}$). In the case of L^1, we have the following dense subsets.

Theorem 15 (Dense subsets of $L^1(\mathbb{R}^d)$)**.** *The following subsets of* $L^1(\mathbb{R}^d)$ *are dense:*

1. *the step functions;*
2. *the simple functions; and*
3. *the (uniformly) continuous functions with compact support.*

In L^1 we have the following invariance properties:

Proposition 6. *Let* $f, g \in L^1(\mathbb{R}^d)$, $h \in \mathbb{R}^d$ *and* $\delta = (\delta_1, \ldots, \delta_n) \in (\mathbb{R}^*)^d$. *Then*

1. *(Additive invariance) The function* $f_h(x) = f(x+h)$ *is integrable, with*
$$\int_{\mathbb{R}^d} f(x+h)dx = \int_{\mathbb{R}^d} f_h = \int_{\mathbb{R}^d} f = \int_{\mathbb{R}^d} f(x)dx$$
2. *(Multiplicative invariance) The function* $x \mapsto f(\delta x)$ *is integrable, with*
$$\int_{\mathbb{R}^d} f(\delta x)dx = \left(\prod_{i=1}^{d} |\delta_i|\right)^{-1} \int_{\mathbb{R}^d} f(x)dx$$
3. *(Convolution is commutative) The map* $y \mapsto f(x-y)g(y)$ *is integrable, with*
$$\int_{\mathbb{R}^d} f(y)g(x-y)dy = \int_{\mathbb{R}^d} f(x-y)g(y)dy$$
4. *(L^1 continuity) We have* $\|f_h - f\|_{L^1} \to 0$ *as* $\|h\| \to 0$.

We further have the following inequality for integrable functions.

Lemma 5 (Tchebychev's inequality)**.** *Let* $f \in L^1(\mathbb{R}^d)$, *and set* $E_\alpha := \{x \in \mathbb{R}^d \mid |f(x)| > \alpha\}$. *Then*

$$m(E_\alpha) \le \frac{\|f\|_{L^1}}{\alpha}$$

3.4 Fubini and Tonelli's theorems

Fubini and Tonelli's theorems concern the integrability and measurability of slice functions, and the interchanging of integrals.

Theorem 16 (Fubini)**.** *Let $f \in L^1(\mathbb{R}^{d_1+d_2})$. Then for almost every $y \in \mathbb{R}^{d_2}$:*

1. *the function f^y defined by $f^y(x) = f(x,y)$ is in $L^1(\mathbb{R}^{d_1})$;*
2. *the function*
$$\left(y \mapsto \int_{\mathbb{R}^{d_1}} f(x,y)dx\right) \in L^1(\mathbb{R}^{d_2})$$
and
3. $$\int_{\mathbb{R}^{d_1+d_2}} f(x,y)dxdy = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)dx\right) dy$$

and the same holds under the interchange $x \leftrightarrow y$.

Tonelli's theorem is the analogous statement with integrable replaced with non-negative and measurable, and proven by taking a sequence of integrable functions tending towards each measurable function.

Theorem 17 (Tonelli)**.** *Let f be non-negative and measurable on $\mathbb{R}^{d_1+d_2}$. Then for almost every $y \in \mathbb{R}^{d_2}$,*

1. *the slice function $f^y(x) = f(x,y)$ is measurable;*
2. *the map $\mathbb{R}^{d_2} \to [0,\infty]$ given by*
$$\left(y \mapsto \int_{\mathbb{R}^{d_1}} f(x,y)dx\right)$$
is measurable; and
3. $$\int_{\mathbb{R}^{d_1+d_2}} f(x,y)dxdy = \int_{\mathbb{R}^{d_2}} \left(\int_{\mathbb{R}^{d_1}} f(x,y)dx\right) dy$$

and the same holds under the interchange $x \leftrightarrow y$.

Fubini's (and more generally Tonelli's) theorem allow us to prove statements about measurability, by taking $f = \chi_E$ for measurable E. We have

Proposition 7. *If $E_1 \times E_2 \subseteq \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ is measurable, and $m_*(E_2) > 0$, then E_1 is measurable.*

Alternatively, if we have E_1 and E_2 measurable, their product is measurable and the measures are compatible:

Proposition 8. *Suppose that $E_1 \subseteq \mathbb{R}^{d_1}, E_2 \subseteq \mathbb{R}^{d_2}$ are measurable. Then $E_1 \times E_2 \subseteq \mathbb{R}^{d_1+d_2}$ is measurable with*
$$m(E_1 \times E_2) = m(E_1)m(E_2)$$
with the understanding that $\infty \cdot 0 = 0$

To prove this we approximate $E_1 \times E_2$ by a uniform product of G_δ sets, and show that their difference is a measure zero set using

Lemma 6. *If $E_1 \subseteq \mathbb{R}^{d_1}, E_2 \subseteq \mathbb{R}^{d_2}$ are measurable, then*

$$m_*(E_1 \times E_2) \leq m(E_1)m(E_2)$$

with the understanding that $\infty \cdot 0 = 0$.

Returning to integrals, we have the following proposition on rewriting an integral:

Proposition 9. *Let $g \in L^1(\mathbb{R}^d)$ and $E_\alpha = \left\{x \in \mathbb{R}^d \mid |g(x)| > \alpha\right\}$. Then*

$$\int_{\mathbb{R}^d} |g(x)|\, dx = \int_0^\infty m(E_\alpha) d\alpha$$

We then have

Corollary 5. *The following hold.*

1. *If $f : \mathbb{R}^{d_1} \to \mathbb{R}$ is measurable, then so is $\left(\tilde{f}(x,y) = f(x)\right) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$;*
2. *if $f \geq 0$ is measurable on $\mathbb{R}^d$, then so is the set*

$$\mathcal{A} = \left\{(x,y) \in \mathbb{R}^d \times \mathbb{R} \mid 0 \leq y \leq f(x)\right\}$$

and

$$m(A) = \int_{\mathbb{R}^d} f(x) dx$$

3. *If f is measurable on $\mathbb{R}^d$, then so is $\tilde{f}(x,y) = f(x-y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$*
4. *(Convolution) If f and g are measurable on $\mathbb{R}^d$, then so is $(x,y) \mapsto f(x-y)g(y)$*

The only non-trivial part is (3), which follows by showing that the preimage of a measure zero set is also measure zero. As a final proposition, we have the actual definition of the Dirichlet convolution of two functions:

Proposition 10. *Let $f, g \in L^1(\mathbb{R}^d)$. Then*

$$(f * g)(x) := \int_{\mathbb{R}^d} f(x-y)g(y)dy$$

is well-defined and finite-valued almost everywhere.

4 Relating derivatives to integrals

4.1 The Lebesgue differentiation theorem

As somewhat of a sanity check that the Lebesgue integration theory up to this point makes sense, we should be able to prove some analogue of the Fundamental Theorem of Calculus (which applies to continuous and hence Riemann integrable functions). The analogue we prove for Lebesgue integrable functions is that

$$\lim_{\substack{m(B) \to 0 \\ x \in B}} \frac{1}{m(B)} \int_B f(y) = f(x) \tag{1}$$

where this limit is taken over balls in $\mathbb{R}^d$ (intuitively shrinking to $\{x\}$). That is, the limit of averages of f taken over balls tending towards x is exactly $f(x)$. We prove this with new tools, ultimately by splitting it into two parts and using a new tool for each. The first such tool is an “upper bound” on a given measurable function, in a way faithful to the expression above.

Definition 13 (Hardy-Littlewood maximal function)**.** *Let f be measurable on $\mathbb{R}^d$. The* Hardy-Littlewood maximal function f^* of f *is*

$$f^*(x) := \sup_{\substack{m(B)\to 0 \\ x\in B}} \frac{1}{m(B)} \int_B |f(y)|\,dy$$

From the definition we see that f^* is non-negative everywhere (being the supremum of a set of non-negative values). We may expect that f^* should be integrable when f is, but this is in fact almost always false.

Proposition 11 (Week 8 workshop question 1)**.** *If $f \in L^1(\mathbb{R}^d)$ is non-zero, then there is $c > 0$ such that*

$$f^*(x) \geq \frac{c}{|x|^d}$$

for all $|x| \geq 1$.

Thus $f^* \notin L^1(\mathbb{R}^d)$ for all functions with $m(\operatorname{supp}(f)) > 0$. We do however have the following properties for the maximal function when f is integrable:

Proposition 12. *Let $f \in L^1(\mathbb{R}^d)$. Then*

1. *f^* is measurable;*
2. *f^* is finite-valued almost everywhere; and*
3. *for $E_\alpha = \left\{x \in \mathbb{R}^d \mid f^*(x)\right\}$, we have $m(E_\alpha) \leq \frac{3^d}{\alpha} \|f\|_{L^1}$*

(3) can be viewed as an analogue of Tchebychev's inequality: even though f^* is non-integrable, it still behaves "similarly" to f in this aspect. To prove (3), we require

Lemma 7 (Vitali covering)**.** *Let $\{B_i\}_{i=1}^n$ be a finite collection of balls. Then there is a subcollection $\left\{B_{i_j}\right\}_{j=1}^k$ of pairwise disjoint balls with*

$$m\left(\bigcup_{i=1}^n B_i\right) \leq 3^d \sum_{j=1}^k m\left(B_{i_j}\right)$$

To show (3) from this, we take an exhaustion of E_α by compact sets, i.e. an increasing sequence of compact sets $\{K_n^\alpha\}_{n=1}^\infty$ with $m(\bigcup_{n=1}^\infty K_n^\alpha \triangle E_\alpha) = 0$ (and in fact $K_n^\alpha \subseteq E_\alpha$, i.e. $K_n^\alpha \nearrow E_\alpha$ up to a set of measure zero), and show that (3) holds for all compact subsets[5] of E_α (and in particular each K_i^α).

Having established these, we note that the constraint in (1) is local, dependent only on balls containing x. To state the corresponding theorem in full generality, we first formalise this "locality".

Definition 14 (Locally integrable, $L^1_{loc}(\mathbb{R}^d)$)**.** *Let f be measurable on $\mathbb{R}^d$. We say that f* is locally integrable *if $f\chi_B$ is integrable for every ball $B \subseteq \mathbb{R}^d$, and write $L^1_{loc}(\mathbb{R}^d)$ for the space of such functions.*

We thus have

Theorem 18 (Lebesgue differentiation)**.** *Let $f \in L^1_{loc}(\mathbb{R}^d)$. Then for almost every $x \in \mathbb{R}^d$,*

$$\lim_{\substack{m(B)\to 0 \\ x\in B}} \frac{1}{m(B)} \int_B f(y)dy = f(x)$$

[5]We can construct the sequence K_n^α by taking $C_n^\alpha \subseteq E_\alpha$ to be such that $m(E_\alpha - C_n^\alpha) < 1/n$, and (by taking unions) so that $C_n^\alpha \subseteq C_m^\alpha$ for $n \leq m$. We can then take $K_n^\alpha = C_n^\alpha \cap \overline{B_n(0)}$, which has $\bigcup_{n=1}^\infty K_n^\alpha = \bigcup_{n=1}^\infty C_n^\alpha$ as $(C_n^\alpha)_{n=1}^\infty$ is increasing.

Applying this to χ_E for a measurable set E, we obtain

Corollary 6 (Points of density)**.** *Let $E \subseteq \mathbb{R}^d$ be measurable. Then*

$$\lim_{\substack{m(B)\to 0\\ x\in B}} \frac{m(B\cap E)}{m(B)} = \chi_E$$

almost everywhere.

That is, at almost every point, this limit is either 1 or 0 (i.e. "binary" in its output). We refer to points which "act like points in E" according to the above corollary by *points of density*:

Definition 15 (Point of density)**.** *Let $E \subseteq \mathbb{R}^d$ be measurable. We say $x \in \mathbb{R}^d$ is a* point of density of *E if*

$$\lim_{\substack{m(B)\to 0\\ x\in B}} \frac{m(B\cap E)}{m(B)} = 1$$

Then **Corollary 6** says that almost every $x \in E$ is a point of density of E, and almost no $x \notin E$ are points of density of E. We also have an associated set for where a slightly stronger variation on the result of the Lebesgue differentiation theorem holds.

Definition 16 (Lebesgue set)**.** *Let f be measurable on $\mathbb{R}^d$. The* Lebesgue set *of f is the set of all x such that*

1. $|f(x)| < \infty$*; and*
2. $\lim_{\substack{m(B)\to 0\\ x\in B}} \int_B |f(y) - f(x)|\, dx = 0$

We then have that

Corollary 7. *Let $f \in L^1_{loc}(\mathbb{R}^d)$. Then almost every $x \in \mathbb{R}^d$ is in the Lebesgue set of f.*

The idea behind the proof is that for each $r \in \mathbb{Q}$ there is a measure zero set E_r with the Lebesgue differentiation theorem holding outside of E_r for $y \mapsto |f(y) - r|$, and so it holds for all such functions (for r ranging over $\mathbb{Q}$) on $E = \bigcup_{r\in\mathbb{Q}} E_r$ of measure zero. Taking $r \in \mathbb{Q}$ with $|f(x) - r| < \varepsilon$ gives

$$\limsup_{\substack{m(B)\to 0\\ x\in B}} \frac{1}{m(B)} \int_B |f(y) - f(x)|\, dy \le \limsup_{\substack{m(B)\to 0\\ x\in B}} \frac{1}{m(B)} \int_B |f(y) - r|\, dy + |f(x) - r| \le 2\varepsilon$$

4.2 Good kernels and approximation to the identity

We consider various families of functions indexed by $\delta \in \mathbb{R}^+$ with nice properties with respect to convergence (as $\delta \to 0$).

Definition 17 (Approximation to the identity)**.** *We say that a family of functions $\{K_\delta\}_{\delta>0}$ is an* approximation to the identity *if for any $f \in L^1(\mathbb{R}^d)$, $(f * K_\delta)(x) \to f(x)$ for almost every $x \in \mathbb{R}^d$.*

That is, the linear maps $f \mapsto f * K_\delta$ "limit" to the identity map $f \mapsto f$ as $\delta \to 0$. We have a slightly weaker definition, which captures the idea that a family of functions "concentrates around zero".

Definition 18 (Good kernel)**.** *We say that $\{K_\delta\}_{\delta>0}$ is a* good kernel *if for all $\delta > 0$,*

1. $\int_{\mathbb{R}^d} K_\delta(x)dx = 1$;
2. $\int_{\mathbb{R}^d} |K_\delta(x)|\, dx \leq A$ *for some* A *independent of* δ*; and*
3. $\lim_{\delta\to 0} \int_{|x|\geq\varepsilon} |K_\delta(x)|\, dx = 0$ *for any* $\varepsilon > 0$.

We have the following sufficient conditions for a family of functions to be an approximation to the identity.

Theorem 19 (Formulation of an approximation to the identity)**.** *Let* $\{K_\delta\}_{\delta>0}$ *be a family of functions satisfying*

1. $\int_{\mathbb{R}^d} K_\delta(x)dx = 1$;
2. $\int_{\mathbb{R}^d} |K_\delta(x)|\, dx \leq A\delta^{-d}$ *for some* A *independent of* δ*; and*
3. $|K_\delta(x)| \leq \frac{A\delta}{|x|^{d+1}}$ *for all* $x \in \mathbb{R}^d$

for each $\delta > 0$*. Then* $\{K_\delta\}_{\delta>0}$ *is an approximation of the identity.*

We show that $(f * K_\delta)(x) \to f(x)$ for almost every $x \in \mathbb{R}^d$ by showing it for x in the Lebesgue set of f (which is sufficient by **Corollary 7**). To show this, we require the following lemma (which we use in showing that $\int_{|y|>\delta} |f(x-y) - f(x)|\, |K_\delta(y)|\, dy \to 0$ as $\delta \to 0$).

Lemma 8. *For* x *in the Lebesgue set of* f*, denote*

$$\mathscr{A}_x(r) = \frac{1}{r^d}\int_{|y|\leq r} |f(x-y) - f(x)|\, dy$$

Then $\mathscr{A}_x$ *is continuous, bounded and* $\lim_{r\to 0} \mathscr{A}_x(r) = 0$.

4.3 The second fundamental theorem of calculus

To prove an analogue of the statement

$$\int_a^b F'(x)dx = F(b) - F(a)$$

we give the following constraint on a function $F : [a, b] \to \mathbb{R}$.

Definition 19 (Absolute continuity)**.** *We say that* $F : [a, b] \to \mathbb{R}$ *is* absolutely continuous *if for any* $\varepsilon > 0$*, there is* $\delta > 0$ *so that if* $\bigsqcup_{n=1}^{\infty} (a_n, b_n) \subseteq [a, b]$ *(where* $a_n < b_n$*) with*

$$\sum_{j=1}^{\infty} (b_n - a_n) < \delta$$

we have

$$\sum_{j=1}^{\infty} |F(b_n) - F(a_n)| < \varepsilon$$

The constraint that a function is absolutely continuous will be key in showing the Fundamental Theorem of Calculus. We also need another notion, namely that the function does not “oscillate too much”.

Definition 20 (Total variation; bounded variation)**.** *Let $F : [a, b] \to \mathbb{R}$ be a function. The* total variation of F *(as a function of x) is*

$$T_F(a, x) = \sup_{\mathcal{P}=\{t_0,\ldots,t_n\}} \sum_{j=1}^{n} |F(t_j) - F(t_{j-1})|$$

where this supremum is taken over partitions $\mathcal{P}$ of $[a, b]$. We say that F is of bounded variation *if $T_F(a, b) < \infty$, and write* $\mathrm{BV}[a, b]$ *for the set of functions on $[a, b]$ of bounded variation.*

Some examples of functions of bounded variation are monotone bounded functions and differentiable functions of bounded derivative, while non-examples include functions of the form $x^a \cos(x^{-b})$ for $0 < b < a$. Note that every function of bounded variation is necessarily bounded (since unboundedness entails unbounded variation). It will be useful to split the variation of a function into the positive and negative variations respectively:

Definition 21 (Positive, negative variation)**.** *Let $F : [a, b] \to \mathbb{R}$. The* positive variation of F *(as a function of x) is*

$$P_F(a, x) = \sup_{\mathcal{P}} \sum_{(+)} (F(t_j) - F(t_{j-1}))$$

and the negative variation of F *is*

$$N_F(a, x) = \sup_{\mathcal{P}} \sum_{(-)} (-(F(t_j) - F(t_{j-1})))$$

where the sum $\sum_{(+)}$ is taken over all j with $F(t_j) \geq F(t_{j-1})$, and $\sum_{(-)}$ is taken over all j with $F(t_j) \leq F(t_{j-1})$.

We can then split the variation of a function F into the positive and negative variations respectively:

Lemma 9. *Let $F : [a, b] \to \mathbb{R}$. Then*

1. *$T_F(a, x) = P_F(a, x) + N_F(a, x)$; and*
2. *$F(x) = F(a) + P_F(a, x) - N_F(a, x)$*

The first result is essentially a reformulation of $\sup(A + B) = \sup(A) + \sup(B)$ for $A, B \subseteq \mathbb{R}$, and the latter involves noting that this equality exact for each individual partition, and that variation increases with refinements (by the triangle inequality). This then gives us the following classification of functions with bounded variation.

Corollary 8. *A function $F : [a, b] \to \mathbb{R}$ is of bounded variation if and only if it is the difference of two bounded, increasing functions.*

The forward implication is (2) in the above lemma and the reverse implication is (1). This then reduces our study of functions of bounded variation to that of bounded, increasing functions, which are simpler. One of the properties of bounded increasing functions is that they have countably many discontinuities (and one way to see this is that there are finitely many discontinuities with jumps bigger than $1/n$ for each $n \in \mathbb{N}$). We have a stronger classification of these bounded increasing functions:

Lemma 10. *Let $F : [a, b] \to \mathbb{R}$ be a bounded, increasing function. Then there is a decomposition $F = F_C + J_F$ where*

1. *$F_C : [a, b] \to \mathbb{R}$ is continuous and increasing; and*

2. *$J_F : [a,b] \to \mathbb{R}$ is ""the" jump function of F", i.e. piecewise constant, and discontinuous exactly where F is, with the same jumps at each point.*

In particular, by the following result J_F is differentiable almost everywhere.

Theorem 20. *Any jump function (increasing, piecewise constant with countably many discontinuities) J_F is differentiable almost everywhere.*

The idea is that since $J_F(x) = \sum_i \alpha_i(x) j_i(x)$ (for j_i corresponding to a discontinuity at x_i with $j_i(x) = 0$ for $x < x_i$, $j_i(x_i) = \theta_i \in [0,1]$ and $j_i(x) = 1$ for $x > x_i$) defines a convergent series, for any $\varepsilon > 0$ the tail terms $J_0(x) = \sum_{i>n} \alpha_i(x) j_i(x)$ of this series should have $\limsup_{h\to 0} \frac{J_0(x+h)-J_0(x)}{h} > \varepsilon$ only in a set of measure zero, by taking a compact subset K (whose measure is larger than $m(E)/2$) and a Vitali covering of K to show that $\eta > m(E)$ for any $\eta > 0$.

We now seem to prove that having bounded variation is sufficient for differentiability almost everywhere, recalling that differentiability of a function F at a point x is the statement that $\lim_{h\to 0} \frac{F(x+h)-F(x)}{h}$ exists (and we refer to this limit as $F'(x)$). By the above lemma, it suffices instead to show this for bounded, continuous, increasing functions. We consider the four Dini numbers (corresponding to infima and suprema of the associated one-sided limits).

Definition 22 (Dini numbers)**.** *Let $F : [a,b] \to \mathbb{R}$ and $x \in (a,b)$. The* Dini numbers *$D^{\pm}F(x)$, $D_{\pm}F(x)$* of *F* at *x are*

1. *(Upper right) $D^+F(x) := \limsup_{h\to 0^+} \frac{F(x+h)-F(x)}{h}$;*
2. *(Upper left) $D^-F(x) := \limsup_{h\to 0^-} \frac{F(x+h)-F(x)}{h}$;*
3. *(Lower right) $D_+F(x) := \liminf_{h\to 0^+} \frac{F(x+h)-F(x)}{h}$; and*
4. *(Lower left) $D_-F(x) := \liminf_{h\to 0^-} \frac{F(x+h)-F(x)}{h}$*

That is, F is differentiable at x if and only if $D^+F(x) < \infty$ and $D^+F(x) = D^-F(x) = D_+F(x) = D_-F(x)$. For F bounded, increasing and continuous we necessarily always have $D_-F(x) \leq D_+F(x) \leq D^-F(x) \leq D^+F(x)$, so it just suffices to show that $D^+F(x) \leq D_-F(x)$ almost everywhere. To show this we use the following lemma:

Lemma 11 (Rising sun)**.** *Suppose that $G : \mathbb{R} \to \mathbb{R}$ is continuous and let*

$$E = \{x \in \mathbb{R} \mid G(x+h) > G(x) \text{ for some } h > 0\}$$

Then E is open, and if E is non-empty we have $E = \bigsqcup_{n=1}^{\infty}(a_k, b_k)$ with $G(a_k) = G(b_k)$ for every finite interval (a_k, b_k).

If instead $G : [a,b] \to \mathbb{R}$, then $G(a_k) \leq G(b_k)$ for an interval with $a_k = a$.

Then to show that $D^+F(x) < \infty$ for F continuous, bounded and increasing, we look at the measurable sets

$$E_r = \left\{x \in [a,b] \mid D^+F(x) \geq r\right\} = \{x \in [a,b] \mid F(x+h) - F(x) > rh \text{ for some } h > 0\}$$

and apply the rising sun lemma to $G(x) := F(x) - rx$, to yield $F(a_k) - ra_k \leq F(b_k) - rb_k$ on each interval (a_k, b_k), and so $m(E_r) = \sum_{k=1}^{\infty}(b_k - a_k) \leq \frac{1}{r}\sum_{k=1}^{\infty}(F(b_k) - F(a_k)) \leq \frac{1}{r}\sum_{k=1}^{\infty}(F(a_{k+1}) - F(a_k)) \leq \frac{1}{r}(F(b) - F(a))$, so $E = \bigcap_{n\in\mathbb{N}} E_{1/n}$ has measure 0.

Now to check that $D^+F(x) \leq D^-F(x)$ for almost every x, we consider the set

$$E = \{x \in [a,b] \mid D^+F(x) > D^-F(x)\} = \bigcup_{R,r\in\mathbb{Q}} \{x \in [a,b] \mid D^+F(x) > R > r > D^-F(x)\}$$

and show that this set necessarily has measure zero (by showing it for any choice of $R, r \in \mathbb{Q}$, since this union is countable). This yields differentiability almost everywhere for bounded, increasing functions, and more generally:

Theorem 21 (Differentiability almost everywhere)**.** *Let $F \in \mathrm{BV}[a,b]$. Then F is differentiable almost everywhere.*

By the Fatou lemma, Lebesgue differentiation theorem and the fact that F' is well-defined almost everywhere for continuous and increasing F, we have a weak version of the fundamental theorem of calculus:

Proposition 13. *Let $F : [a,b] \to \mathbb{R}$ be a continuous, increasing function. Then*

$$\int_a^b F'(x)dx \leq F(b) - F(a)$$

To show the second fundamental theorem of calculus, we require a lemma which says that given a nice covering (in the sense that points can be approximated arbitrarily well) of E, we can choose balls which cover most of E, in the sense that the part of E not covered by the balls can be made arbitrarily small.

Definition 23 (Vitali covering)**.** *A collection of balls $\mathcal{B} = \{B_i\}_{i\in I}$ is called a Vitali covering of a set E if for any $x \in E$ and $\eta > 0$, there is a ball $B_{x,\eta} \in \mathcal{B}$ with $x \in B_{x,\eta}$ and $m(B_{x,\eta}) < \varepsilon$.*

We then have the following result.

Lemma 12 (Vitali's 2nd covering lemma)**.** *Let E be a set of finite measure with a Vitali covering $\mathcal{B}$. Then for any $\delta > 0$, there is a finite subcollection $\{B_i\}_{i=1}^n$ of pairwise disjoint balls in $\mathcal{B}$ so that*

$$m\left(\bigsqcup_{i=1}^n B_i\right) = \sum_{i=1}^n m(B_i) \geq m(E) - \delta$$

We thus obtain the following result, which also gives a classification of which functions arise as the integrals of integrable functions.

Theorem 22 (The second fundamental theorem of calculus)**.** *If $F : [a,b] \to \mathbb{R}$ is absolutely continuous, then $F'(x)$ exists for almost every x, and is integrable with*

$$F(x) = \int_a^x F'(t)dt + F(a)$$

Conversely, if f is integrable, then

$$F(x) = \int_a^x f(t)dt$$

is absolutely continuous with $F'(x) = f(x)$.

The above result gives us a finer decomposition of bounded increasing functions:

Corollary 9. *Let $F : [a,b] \to \mathbb{R}$ be a bounded, increasing function. Then there is a decomposition*

$$F = F_A + F_C + J_F$$

where all of F_A, F_C, J_F are increasing, and

1. *F_A is absolutely continuous;*
2. *F_C has vanishing derivative almost everywhere; and*
3. *J_F is a jump function (i.e. piecewise constant).*

5 Hilbert spaces

We generalise the notions of finite-dimensional inner product spaces to infinite (countable)-dimensional inner product spaces. This comes with a completion issue, in the sense that the naive limit

$$\bigcup_{k=1}^{\infty} \mathbb{C}^k \subseteq \mathbb{C}^\infty$$

of sequences which are eventually zero is not complete with the induced inner product, so we need to assert this separately, or equivalently take a "closure" in some corresponding complete ambient space.

Definition 24 (Hilbert space)**.** *A* Hilbert space *H is a complete (Hermitian) inner product space.*

Remark 3. *In topology, we say a topological space is* separable *if it admits a countable dense subset. In the case of a Hilbert space H, this is equivalent to the existence of a countable set $\{v_i\}_{i=1}^\infty$ whose span is dense in H (by taking all finite $\mathbb{Q}$ or $\mathbb{Q}[i]$ linear combinations of these vectors).*

We take all Hilbert spaces to be separable unless otherwise stated. Naturally, as a Hilbert space is an inner product space, we have the following properties.

Remark 4. *Let H be a Hilbert space, and $x, y \in H$. Then*

1. *(Induced norm) $\|x\| := \langle x, x\rangle^{1/2}$ is a norm on H;*
2. *(Cauchy-Schwarz inequality) $|\langle x, y\rangle| \le \|x\| \, \|y\|$; and*
3. *(The parallelogram law) $\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$.*

In fact, a norm $\|\cdot\|$ satisfies (3) if and only if it is induced by some inner product, namely that given by

$$\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$$

for a real inner product, and

$$\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2) + \frac{i}{4}(\|x/i+y\|^2 - \|x/i-y\|^2)$$

for a complex Hermitian inner product.

For a somewhat natural infinite-dimensional Hilbert space, we can consider the space of square-integrable functions:

Definition 25 ($L^2(\mathbb{R}^d)$)**.** *For any $d \in \mathbb{N}$, the* space $L^2(\mathbb{R}^d)$ of square-integrable functions *is*

$$L^2(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{C} \;\middle|\; \int_{\mathbb{R}^d} |f|^2 < \infty \right\} / \sim$$

where $f \sim g$ if and only if $f = g$ almost everywhere, with the L^2 inner product

$$\langle f, g \rangle_{L^2(\mathbb{R}^d)} := \int_{\mathbb{R}^d} f\overline{g}$$

We have the following properties, showing that $L^2(\mathbb{R}^d)$ is indeed a separable Hilbert space:

Theorem 23. *The following hold for $L^2(\mathbb{R}^d)$.*

1. *$L^2(\mathbb{R}^d)$ is complete in the L^2-norm;*
2. *$L^2(\mathbb{R}^d)$ is separable, that is, admits a countable subset of dense span.*

The idea behind (1) is similar to that of showing L^1 is complete, and for (2) we can take the indicator functions of rational rectangles. As in the finite-dimensional case, it is easier to study a Hilbert space through the reference frame of an orthonormal basis. We redescribe what this means for an infinite-dimensional space, in a way which allows for infinite linear combinations.

Definition 26 (Orthonormal set; orthonormal basis)**.** *A set $\{e_i\}_{i=1}^{\infty}$ in a Hilbert space H is an* orthonormal set *if for any $i, j \in \mathbb{N}$,*

$$\langle e_i, e_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

We say that $\{e_i\}_{i=1}^{\infty}$ is an orthonormal basis *for H if $H = \overline{\operatorname{span}\{e_i\}_{i=1}^{\infty}}$, that is, $\{e_i\}_{i=1}^{\infty}$ has dense span in H.*

With respect to an orthonormal set, the following relations hold with norms.

Proposition 14 (Bessel's inequality)**.** *Let $f \in H$ and $\{e_i\}_{i=1}^{\infty}$ be an orthonormal set. Then*

$$\|f\|^2 \geq \sum_{k=1}^{\infty} |\langle f, e_k \rangle|^2$$

with equality iff f is in the closure of $\operatorname{span}\{e_i\}_{i=1}^{\infty}$.

We also have the following classification of when orthonormal sets are orthonormal bases:

Theorem 24. *The following are equivalent for an orthonormal set $\{e_k\}_{k=1}^{\infty}$:*

1. *$\{e_k\}_{k=1}^{\infty}$ is an orthonormal basis;*
2. *if $\langle f, e_k \rangle = 0$ for all k, then $f = 0$;*
3. *the partial sums $S_n(f) = \sum_{k=1}^{n} \langle f, e_k \rangle e_k$ converge to f.*
4. *for any $f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k$, $\|f\|^2 = \sum_{k=1}^{\infty} |\langle f, e_k \rangle|^2$*

By applying an infinite version of the Gram-Schmidt process, we obtain the following result:

Theorem 25. *Any separable Hilbert space H admits a countable orthonormal basis.*

We have a notion of equivalence between Hilbert spaces (and more generally inner product spaces), in the sense that there is a map $T : H_1 \to H_2$ which preserves inner products (i.e. $\langle Tf, Tg \rangle_{H_2} = \langle f, g \rangle_{H_1}$), which we refer to as *unitary equivalences* or *isometries*. Based on the previous theorem, we have

Corollary 10. *Every separable Hilbert space is isometric to the space of square-summable sequences $\ell^2(\mathbb{N})$.*

5.1 Orthogonal projections

Given a subspace $S \subseteq H$, we have $S^\perp = \{w \in H \mid \langle v, w \rangle = 0 \text{ for all } v \in S\}$, its usual orthogonal complement. Unlike the finite-dimensional case, we do not always have a decomposition $H = S \oplus S^\perp$, and the key insight behind this is that $S^\perp = \langle S, - \rangle^{-1}(0) = \bigcap_{v \in S} \langle v, - \rangle^{-1}(0)$ is always closed, alongside the fact that $S^\perp = (\overline{S})^\perp$. We instead actually have $(S^\perp)^\perp = \overline{S}$, which suggests the following result.

Lemma 13. *Let $S \subseteq H$ be a closed subspace. Then*

$$H = S \oplus S^\perp$$

and for any $f \in H$ there is a unique $g_0 \in S$ with $f - g_0 \in S^\perp$.

We thus have an orthogonal projection map $P_S : H \to H$ obtained by writing $w \in H$ as $w_1 + w_2$ (for $w_1 \in S$, $w_2 \in S^\perp$), and this satisfies $\mathrm{id}_H = P_S + P_{S^\perp}$ and $\|P_S f\| \leq \|f\|$ for all $f \in H$.

The following properties also hold for orthogonal projections:

1. $S^\perp = \overline{S}^\perp$, and in particular $(S^\perp)^\perp = \overline{S}$;
2. In particular, $S^\perp = T$ implies $S = T^\perp$ if and only if S is closed.

We also have the following examples of Hilbert spaces and some natural associated orthonormal bases:

Example 1. *1. Any finite-dimensional Euclidean space $\mathbb{R}^n$ or $\mathbb{C}^n$;*

2. The space $L^2[-\pi, \pi]$ of square-integrable functions on $[-\pi, \pi]$, with inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f \overline{g}$$

orthonormal basis given by the Fourier basis $\left\{e_n : t \mapsto e^{int}\right\}_{n \in \mathbb{Z}}$;

3. The space $\ell^2(\mathbb{Z}) = \left\{(a_n)_{n \in \mathbb{Z}} \mid \sum_{n \in \mathbb{Z}} |a_n|^2\right\}$ of square summable sequences, with orthonormal basis $\{e_i = (\delta_{ij})_{j \in \mathbb{Z}}\}$;

4. The (non-separable) space[6] $\ell^2(\mathbb{R}^d) = \left\{f : \mathbb{R}^d \to \mathbb{C} \mid \sum_{x \in \mathbb{R}^d} |f(x)|^2 < \infty\right\}$ of square-summable functions on $\mathbb{R}^d$, with (uncountable) orthonormal basis $\left\{\chi_{\{x\}}\right\}_{x \in \mathbb{R}^d}$.

5.2 Bounded operators

We can also consider the space of linear transformations between two Hilbert spaces H_1 and H_2, which is naturally a vector space under pointwise operations. To describe a norm on such transformations, we first describe what it means for such an operator to be "bounded".

Definition 27. *We say that $T : H_1 \to H_2$ is* bounded *if there is $M > 0$ such that*

$$\|Tf\|_{H_2} \leq M \|f\|_{H_1}$$

for all $f \in H_1$. We write $B(H_1, H_2)$ for the space of bounded transformations $H_1 \to H_2$.

We then have a reasonable notion of "how large" an operator is.

[6] Note that $f^{-1}(\mathbb{C} \setminus \{0\})$ is necessarily countable in this case.

Definition 28 (Operator norm)**.** *Let* $T \in B(H_1, H_2)$. *Its* operator norm $\|T\|$ *is*

$$\|T\| = \inf_{M>0} \left\{ \|Tf\|_{H_2} \leq M \|f\|_{H_1} \right\}$$

Remark 5. *A more useful (or rather explicit) way to think about this is that* $\|T\|$ *is such that*

$$\|Tf\|_{H_2} \leq \|T\| \|f\|_{H_1}$$

and that it is minimal with this property (in the sense that for any M *with* $\|Tf\|_{H_2} \leq M \|f\|_{H_1}$*, we have* $\|T\| \leq M$*).*

We have the following equivalent formulations of the operator norm.

Lemma 14. *Let* $T \in B(H_1, H_2)$. *Then*

$$\|T\| = \sup_{\|f\|_{H_1}, \|g\|_{H_2} \leq 1} \left| \langle Tf, g \rangle_{H_2} \right| = \sup_{f \in H_1, f \neq 0} \frac{\|Tf\|_{H_2}}{\|f\|_{H_1}}$$

The suprema above can be replaced with ranging over the respective unit spheres in H_1 and H_2. Bounded operators relate to continuous functionals in the following way:

Proposition 15. *Let* $T : H_1 \to H_2$ *be an operator. Then* T *is bounded if and only if it is continuous.*

5.3 The Riesz representation theorem

For a Hilbert space H, we can consider the dual space $\mathrm{Hom}(H, \mathbb{C})$, though for our purposes we want to consider the corresponding operator norms, so it will be easier to consider the space of continuous functionals $H^* = B(H, \mathbb{C})$. We have the following identification of H with its dual space H^* in this case.

Theorem 26 (Riesz representation)**.** *The map*

$$\begin{aligned} \ell : H &\to H^* \\ f &\mapsto (g \mapsto \langle g, f \rangle) \end{aligned}$$

is a linear isometry $H \cong H^*$.

As such, given a bounded operator T, we have

Definition 29 (Adjoint operator)**.** *Let* $T \in B(H_1, H_2)$. *Its* adjoint operator $T^* \in B(H_2, H_1)$ *is such that*

$$\langle Tf, g \rangle_{H_2} = \langle f, T^* g \rangle_{H_1}$$

and by the above theorem, this is well-defined:

Proposition 16. *Let* $T \in B(H_1, H_2)$. *Then there is a unique* $T \in B(H_2, H_1)$ *with the above property, and* $\|T^*\| = \|T\|$ *and* $T^{**} = T$.

With adjoints, we have the following classification of orthogonal projections:

Proposition 17. *An operator* $P : H \to H$ *is orthogonal projection if and only if* $P^2 = P$ *and* $P^* = P$

That is, an operator is an orthogonal projection exactly when it is a self-adjoint projection. We also have the following properties of the adjoint:

1. $(S+T)^* = S^* + T^*$;
2. $(ST)^* = T^* S^*$;
3. $(\lambda S)^* = \overline{\lambda} S^*$;
4. if T is an isometry, then $T^* T = I$

In relation to orthogonal projections, we have

1. $\ker(T) = \mathrm{range}(T^*)^\perp$ and $\ker(T^*) = \mathrm{range}(T^*)^\perp$; and
2. if $\mathrm{range}(T^*)$ is closed, $\ker(T)^\perp = \mathrm{range}(T^*)$

5.4 Integral operators, Hilbert-Schmidt and trace-class operators

On $L^2(\mathbb{R}^d)$ we have a class of operators given by integrating a fixed K on $\mathbb{R}^d \times \mathbb{R}^d$ against an input f:

Definition 30. *Let $K \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$. The* integral operator T_K with kernel K *is*

$$(T_K f)(x) := \int_{\mathbb{R}^d} K(x,y) f(y) dy$$

For $K \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$, this is well-defined (but may not be for general measurable K):

Proposition 18. *Let T be an integral operator with kernel $K \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$. Then*

1. *For any $f \in L^2(\mathbb{R}^d)$ and almost every $x \in \mathbb{R}^d$, the map*

$$y \mapsto K(x,y) f(y)$$

   is integrable;
2. *$T_K \in L^2(\mathbb{R}^d)$ with $\|T_K\|_{L^2(\mathbb{R}^d)} \le \|K\|_{L^2(\mathbb{R}^{2d})}$; and*
3. *T_K^* is also an integral operator with kernel $(x,y) \mapsto \overline{K(y,x)}$.*

We have another class of operators, given by being "square summable" with respect to some orthonormal basis.

Definition 31 (Hilbert-Schmidt)**.** *We say that $T : H \to H$ is a* Hilbert-Schmidt operator *if there is an orthonormal basis $\{e_n\}_{n=1}^\infty$ for H with*

$$\sum_{n=1}^\infty \|Te_n\|^2 < \infty$$

and define the square of its Hilbert-Schmidt norm $\|T\|_{HS}$ *to be the above sum of squares.*

Every integral operator T_K is Hilbert-Schmidt with the following compatibility of norms:

Lemma 15. *For $K \in L^2(\mathbb{R}^d \times \mathbb{R}^d)$, we have $\|T_K\|_{HS} = \|K\|_{L^2(\mathbb{R}^d \times \mathbb{R}^d)}$.*

We also have a class of operators for which we can define a reasonable notion of a "trace":

Definition 32 (Trace class operator)**.** *We say that $T : H \to H$ is* trace-class *if there is an orthonormal basis $\{e_n\}_{n=1}^\infty$ with*

$$\mathrm{tr}(T) := \sum_{n=1}^\infty \|Te_n\| < \infty$$

5.5 Special types of operators

We have various classes of operators, which we summarise below.

Definition 33 (Finite-rank; symmetric; diagonalisable;)**.** *Let $T : H \to H$ be an operator. We say that T is*

1. finite-rank *if $T(H)$ is finite-dimensional;*
2. symmetric *if $T^* = T$;*
3. diagonalisable *if it admits an orthonormal eigenbasis $\{\phi_k\}_{k=1}^{\infty}$; and*
4. compact *if one (and hence all) of*
   (a) *$\overline{T(B)}$ is compact where $B \subseteq H$ is the closed unit ball;*
   (b) *$\overline{T(S)}$ is compact for any bounded $S \subseteq H$; or*
   (c) *For every sequence $(f_n)_{n\in\mathbb{N}}$, $(Tf_n)_{n\in\mathbb{N}}$ has a convergent subsequence.*

Remark 6. *We can equivalently view these classes of operators in the following senses:*

Bounded : Maps bounded sets to bounded sets;
Compact : Maps bounded sets to compact sets;
Diagonalisable : Given by scaling components in some choice of coordinates

In regards to compact operators, the following properties also hold:

1. If S is bounded and T is compact, then ST and TS are compact;
2. if (T_n) is a sequence of compact operators with $\|T_n - T\| \to 0$, then T is also compact;
3. if T is compact, then there is a sequence (T_n) of operators of finite rank with $\|T_n - T\| \to 0$; and
4. T is compact if and only if T^* is compact

5.6 The spectral theorem

In some sense, the compact operators are those which most closely resemble those on finite-dimensional space, and we have a spectral theorem for these operators (in analogy to the finite-dimensional case). We begin by describing various lemmas on compact and symmetric operators.

Lemma 16. *Let T be diagonalisable with respect to $\{\phi_k\}_{k\in\mathbb{N}}$ with $T\phi_k = \lambda_k\phi_k$. Then T is compact if and only if $\lambda_k \to 0$.*

For symmetric operators, we have the following restrictions on eigenvectors and eigenvalues.

Lemma 17. *Let T be symmetric. Then*

1. *Every eigenvalue of T is real; and*
2. *if ϕ_1, ϕ_2 are eigenvectors with differing eigenvalues, then they are orthogonal.*

For compact operators we also have the following dimensionality statement for the eigenspaces:

Lemma 18. *Let T be compact. Then*

1. *For any non-zero eigenvalue λ, the eigenspace $\ker(T - \lambda I)$ is finite-dimensional;*
2. *T has countably many eigenvalues $\{\lambda_k\}_{k=1}^{\infty}$, with $\lambda_k \to 0$ as $k \to \infty$; and*

3. *For any $\mu > 0$, the direct sum of the eigenspaces*

$$E_\mu = \operatorname{span}_{\mathbb{C}} \{f \mid Tf = \lambda f, |\lambda| > \mu\} = \bigoplus_{|\lambda|>\mu} \ker(T - \lambda I)$$

is finite-dimensional.

We have the following key ingredient for non-zero, compact, symmetric operators:

Lemma 19. *Let $T \neq 0$ be compact and symmetric. Then either $\|T\|$ or $-\|T\|$ is an eigenvalue of T.*

With these, we have the spectral theorem in (countably) infinite dimensions:

Theorem 27. *Let T be an operator from a Hilbert space H to itself. The following are equivalent.*

1. *T is compact and symmetric;*
2. *T is diagonalisable with orthonormal eigenbasis $\{\phi_k\}_{k=1}^{\infty}$ with $T\phi_k = \lambda_k \phi_k$, $\lambda_k \in \mathbb{R}$ for all k, and $\lambda_k \to 0$ as $k \to \infty$.*