INTEGRATION: H.T. 2025, 16 LECTURES

Acknowledgement. These notes are a very small edit of the notes produced by Charles Batty who lectured this course from 2018-21. I'm grateful to him for allowing me to use his notes in this way. I am responsible for any typos / inaccuracies in the notes (please let me know of any you find).

Stuart White
`stuart.white@maths.ox.ac.uk`

Contents

Reading

Z. Qian, *Part A: Integration*, Available on course moodle page.
M. Capinski and E. Kopp, *Measure, Integration and Probability*, Springer SUMS (2nd edition, 2004)
H.A. Priestley, *Introduction to Integration*, OUP, 1997
S. Axler. *Measure Theory, Integration & Real Analysis*, Springer, Graduate Texts in Mathematics, 2020. This book is open access (`https://measure.axler.net/`)
E. M. Stein & R. Shakarchi, *Real Analysis: Measure Theory, Integration and Hilbert Spaces*, Princeton Lectures in Analysis III, Princeton University Press, 2005

Date: Version from 13 Feb 2025.

D.J.H. Garling, *A Course in Mathematical Analysis, III (Part 6)*, CUP, 2014.

Qian's notes were written for the course as he gave it in 2014-17, based on previous versions of the course given by Alison Etheridge and Charles Batty. We will cover more or less the same material, but not follow his notes exactly.

Capinski and Kopp is the most basic of the books, giving the theory in a basic style, but with not many worked examples; we shall follow rather closely their approach to the theory. Priestley adopts a very different approach to the construction of the integral, so early parts of her book look quite different from what we will do, but about the 8th lecture onward everything comes together; she has lots of worked examples.

Stein and Shakarchi, and Garling, are a little more sophisticated in the theory. Garling's book is based on lectures given in Cambridge, and it has a good number of worked examples.

Numerous other useful books may be found in libraries. Some may adopt different approaches to the construction of the integral, but when they talk about Lebesgue integration they all mean the same class of integrable functions and the same theorems.

Introduction

In Prelims, you saw how to define $\int_a^b f(x)\,dx$ for a continuous function $f : [a,b] \to \mathbb{R}$ or more generally for Riemann integrable f. It had some good properties: the Fundamental Theorem of Calculus shows that it is more or less an inverse of differentiation, leading to rigorous statements concerning A level calculus. Moreover you saw that

$$(*) \qquad \lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx$$

if (f_n) converges to f uniformly on $[a,b]$. This was useful (a) for integrating power series term-by-term, (b) for finding $\lim_{n\to\infty} \int_\gamma f_n(z)\,dz$, where γ is a contour of finite length, in complex analysis last term. However, the Riemann integral has various deficiencies:

(a) There are still functions which one feels one should be able to integrate, for which the Prelims definition fails to work. For example, let $f = \chi_{\mathbb{Q}\cap[0,1]}$ be the characteristic function of $\mathbb{Q} \cap [0,1]$. Then

$$\underline{\int_0^1} f(x)\,dx = 0, \qquad \overline{\int_0^1} f(x)\,dx = 1$$

so the definition of the integral fails.

In particular, if we want to define the length of a subset E of $\mathbb{R}$ by

$$m(E) = \int \chi_E(x)\,dx,$$

we need to extend the definition of integrals in some way beyond Riemann integration.

(b) There is a lack of theorems saying that

$$f_n \to f \implies \int f_n(x)\,dx \to \int f(x)\,dx$$

particularly for integrals over $\mathbb{R}$ or unbounded subsets of $\mathbb{R}$. To some extent, this is unavoidable because of the following example:

Example 0.1. Let $f_n(x) = n^2x^n(1-x)$ $(0 \le x \le 1)$. Then $\lim_{n\to\infty} f_n(x) = 0$ for all $x \in [0,1]$, but $\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1$.

This example is going to arise in any reasonable theory. But we would like some more theorems of the form

> Suppose (f_n) is a sequence of integrable functions, $f_n(x) \to f(x)$ for each x, and [*supplementary assumptions to be inserted*]. Then f is integrable and $\int f(x)\,dx = \lim_{n\to\infty} \int f_n(x)\,dx$.

Lebesgue's integration theory provides two very powerful theorems of this form (Monotone Convergence Theorem, Dominated Convergence Theorem). The theorems are less good in Riemann integration, because one has to assume that the limiting function is integrable.

(c) Riemann's integration theory does not generalise to include various other contexts such as:
  - probability theory, taking expectations of arbitrary random variables (continuous, discrete, hybrid, singular);
  - summing infinite series.

Lebesgue's theory resolves these difficulties, except where there is an unavoidable obstruction. In a sense the passage from Riemann integration to Lebesgue integration resembles the passage from rational numbers to real numbers—it completes the space of integrable functions, or it fills in the gaps.

The crucial ideas of the Lebesgue's construction are:

(i) Instead of using integrals to define lengths of sets, define the length of a set directly; then define integrals.
(ii) Instead of partitioning the x-axis into intervals and using step functions, partition the y-axis into intervals and considering corresponding "simple" functions.

There are other ways of constructing Lebesgue's integral on $\mathbb{R}$, including ways which use step functions (see Priestley), but they don't generalise so easily to probability (for example). Once one gets the Monotone Convergence Theorem, then everything is the same, however you got there. We then get a whole host of theorems about:

- passing limits through integrals,
- passing infinite sums through integrals,
- differentiating through integrals,
- interchanging two integrals (Fubini's Theorem)
- changing variables.

Note that these processes do not always work—there are simple counterexamples for the first 4! So all these theorems have conditions which must be checked before using in applications. In this course, we do not take the position that you can just assume all these processes work. On the other hand, we shall not go pedantically through all details of the construction of the integral and the proofs of the theorems.

We'll approach the construction in a way which generalises easily, but the proofs of these generalistions are often not interesting. The construction up to the MCT will take some time - around 8 lectures - and then useful theorems and applications will come thick and fast.

Please be aware that all the Prelims theory remains valid in this context. Lebesgue integration theory extends Riemann's theory by enabling you to integrate more functions. In particular, the Fundamental Theorem of Calculus (both versions), Integration by Parts and Substitution remain valid under the assumptions given in Prelims.

1. Extended real number system

In this course, we shall often take infinite series of non-negative terms and limits of (monotone) sequences. In order to avoid complications concerning divergence, it will be convenient to work in the extended real numbers including $-\infty$ and ∞, and to use the notions of $\limsup$ and $\liminf$.

Thus we consider the set $[-\infty,\infty] = \mathbb{R} \cup \{-\infty,\infty\}$. Addition and multiplication by ∞ are defined as follows (for $x \in \mathbb{R}$):

$$\begin{aligned} x + \infty &= \infty + x = \infty, \\ x - \infty &= -\infty + x = -\infty, \\ x.\infty &= \infty.x = (-x).(-\infty) = \begin{cases} \infty & (x > 0), \\ -\infty & (x < 0), \\ 0 & (x = 0). \end{cases} \end{aligned}$$

Note that

- $\infty - \infty$ is undefined;
- the usual laws (commutativity, associativity and distributivity) apply, provided that the relevant expressions are defined;
- the above are uncontroversial, except for $0.\infty = 0$ which is convenient for our particular context but might be inappropriate in other mathematical contexts.

The ordering on $[-\infty,\infty]$ is the obvious one, and $\lim_{n\to\infty} a_n = \infty$ has the same meaning as in Prelims Analysis.

In this system, any subset E has a supremum and an infimum in $[-\infty,\infty]$. Note that $\sup \emptyset = -\infty$. If $E \subseteq \mathbb{R}$, $\sup E = \infty$ if and only if E is not bounded above. For an increasing sequence (a_n), $\lim_{n\to\infty} a_n = \sup\{a_n\}$. If $a_n \geq 0$ for all n, then $\sum a_n = \infty$ if and only if the series diverges.

Proposition 1.1. 1. *Let (a_n) be a sequence of non-negative terms. Then*

$$\sum_{n=1}^{\infty} a_n = \sup \left\{ \sum_{n\in J} a_n : J \text{ finite subset of } \mathbb{N} \right\}.$$

2. *Let* $(b_{mn})_{m,n\geq 1}$ *be a double sequence of non-negative terms, and* $\{(m_k,n_k) : k \geq 1\}$ *be any enumeration of* $\mathbb{N}\times\mathbb{N}$. *Then*

$$\sum_{m=1}^{\infty}\sum_{n=1}^{\infty} b_{mn} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} b_{mn} = \sum_{k=1}^{\infty} b_{m_k,n_k} = \sup\left\{\sum_{(m,n)\in J} b_{mn} : J \textit{ finite subset of } \mathbb{N}\times\mathbb{N}\right\}.$$

In particular, Proposition 1.1 implies that $\sum a_n$ is independent of the order of the terms, and similarly $\sum\sum b_{mn}$ can be arbitrarily rearranged.

A bounded sequence (a_n) in $\mathbb{R}$ may not have a limit. It has a supremum and infimum, but for some large values of n, a_n may not be close to them. Think for example about $a_n = (1 + 1/n)\sin n$. Asymptotically the values oscillate between -1 and 1, but there are infinitely many values bigger than 1 and infinitely many smaller than -1.

For a sequence (a_n) in $[-\infty,\infty]$, define

$$\limsup_{n\to\infty} a_n = \lim_{m\to\infty}\left(\sup_{n\geq m} a_n\right),$$
$$\liminf_{n\to\infty} a_n = \lim_{m\to\infty}\left(\inf_{n\geq m} a_n\right).$$

The limits exist, because $\left(\sup_{n\geq m} a_n\right)_{m\geq 1}$ is a decreasing sequence

So, $\limsup_{n\to\infty} a_n$ is the largest number ℓ such that there is a subsequence of (a_n) converging to ℓ.

Examples 1.2. 1. Let $a_n = (1 + 1/n)\sin n$. Then

$$\limsup_{n\to\infty} a_n = 1, \qquad \liminf_{n\to\infty} a_n = -1.$$

2. Let $a_n = (-1)^n$. Then

$$\limsup_{n\to\infty} a_n = 1, \qquad \liminf_{n\to\infty} a_n = -1.$$

3. Let $a_n = n(-1)^n$. Then

$$\limsup_{n\to\infty} a_n = \infty, \qquad \liminf_{n\to\infty} a_n = -\infty.$$

4. Let $a_n = \begin{cases} 1 + 2^{-n} & (n \text{ prime}), \\ 0 & \text{otherwise.}\end{cases}$ Then

$$\limsup_{n\to\infty} a_n = 1, \qquad \liminf_{n\to\infty} a_n = 0.$$

Proposition 1.3. 1. $\liminf_{n\to\infty} a_n = -\limsup_{n\to\infty}(-a_n)$*;*
2. $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$*;*
3. $\lim_{n\to\infty} a_n$ *exists if and only if* $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$*; then all are equal;*
4. *If* $a_n \leq b_n$ *for all* n*, then* $\limsup_{n\to\infty} a_n \leq \limsup_{n\to\infty} b_n$*;*
5. $\limsup_{n\to\infty}(a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$ *(if all sums exist).*

lim sup and lim inf are useful for avoiding epsilontics. For example, consider the Sandwich Rule, i.e., suppose that $a_n \le b_n \le c_n$ for all n and $\lim a_n = \lim c_n$. Then

$$\begin{aligned}
\limsup b_n &\le \limsup c_n \qquad \text{(Proposition 1.3(4))}\\
&= \lim c_n \qquad \text{(Proposition 1.3(3))}\\
&= \lim a_n \qquad \text{(assumption)}\\
&= \liminf a_n \qquad \text{(Proposition 1.3(3))}\\
&\le \liminf b_n \qquad \text{(Proposition 1.3(4))}\\
&\le \limsup b_n \qquad \text{(Proposition 1.3(2))}.
\end{aligned}$$

Hence equality holds throughout, so $\lim b_n = \lim a_n$, by Proposition 1.3(3).

2. Lebesgue measure

A measure of length for (all) subsets of $\mathbb{R}$ should be a function $m : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ satisfying:

(i) $m(\emptyset) = 0$, $m(\{x\}) = 0$;
(ii) $m(I) = b - a$ if I is an interval with endpoints a, b, where $a < b$;
(iii) $m(A + x) = m(A)$;
(iv) $m(\alpha A) = |\alpha| m(A)$;
(v) $m(A) \le m(B)$ if $A \subseteq B$; (m is *monotone*);
(vi) $m(A \cup B) = m(A) + m(B)$ if $A \cap B = \emptyset$ (m is *finitely additive*);
(vi)′ $m\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty m(A_n)$ if $A_n \cap A_k = \emptyset$ for $k \ne n$ (m is *countably additive*);
(vii) $m\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} m(A_n)$ if (A_n) is an increasing sequence of sets.

In fact, there is very considerable redundancy here. For example, (v), (vi) and (vii) follow from (i) and (vi)′.

The status of (vi)′ is perhaps debatable, but it is usually assumed. It is equivalent to (vi) and (vii) together, and (vii) is essential to have a Monotone Convergence Theorem.

Let us attempt to construct such an m. For $A \subseteq \mathbb{R}$, suppose that $A \subseteq \bigcup_{n=1}^\infty I_n$ for intervals I_n. Letting $I_n' = I_n \setminus (I_1 \cup \cdots \cup I_{n-1})$, we have

$$m(A) \le m\left(\bigcup I_n'\right) = \sum m(I_n') \le \sum m(I_n).$$

So we attempt to define m as follows. First, for any interval I with endpoints a and b, define

$$|I| = b - a.$$

For $A \subseteq \mathbb{R}$, we define the *outer measure* of A to be

$$m^*(A) = \inf \left\{ \sum_{n=1}^\infty |I_n| : I_n \text{ intervals}, A \subseteq \bigcup_{n=1}^\infty I_n \right\}.$$

We can always take $I_n = [-n, n]$, so the infimum is not over the empty set (but $m^*(A)$ may be infinite). It makes no difference if we restrict I_n to being closed intervals, or open intervals.

Proposition 2.1. 1. *$m^*(\emptyset) = 0$, $m^*(\{x\}) = 0$;*
2. *$m^*(I) = |I| = b - a$ if I is any interval with endpoints a, b;*

3. $m^*(A+x) = m^*(A)$;
4. $m^*(\alpha A) = |\alpha| m^*(A)$;
5. $m^*(A) \le m^*(B)$ *if* $A \subseteq B$;
6. $m^*(A \cup B) \le m^*(A) + m^*(B)$;

6′. $m^*\left(\bigcup_{n=1}^\infty A_n\right) \le \sum_{n=1}^\infty m^*(A_n)$.

Proof. (1), (3), (4), (5) are easy; (6) and (6)' are moderately tricky exercises. See Q8 Sheet 1. Let us prove (2); we will do it for $I = [a,b]$; then the other cases follow using (1), (5) and (6).

Firstly, $m^*[a,b] \le b-a$, because we may take $I_1 = [a,b]$ and $I_n = \{0\}$ for $n \ge 2$.

Now suppose that $[a,b] \subseteq \bigcup_{n=1}^\infty I_n$ where I_n is an interval with endpoints a_n, b_n (which we can assume interesects $[a,b]$). Take $\varepsilon > 0$. Let

$$J_n = \left(a_n - \varepsilon 2^{-n}, b_n + \varepsilon 2^{-n}\right) =: (c_n, d_n).$$

Then J_n is open and $[a,b] \subseteq \bigcup_{n=1}^\infty J_n$. By the Heine-Borel Theorem, $[a,b]$ is compact, so $[a,b] \subseteq \bigcup_{n=1}^N J_n$ for some N.

Now it is almost obvious that $b-a \le \sum_{n=1}^N |J_n|$. Enumerate $\{c_n, d_n : n = 1, \dots, N\}$ in increasing order:

$$x_1 < x_2 < \cdots < x_k.$$

Then $x_1 < a < b < x_k$, each interval (x_i, x_{i+1}) is contained in some J_n, and J_n has endpoints $c_n = x_{k_n}$, $d_n = x_{\ell_n}$, say. Hence

$$b - a < x_k - x_1 = \sum_{i=1}^{k-1} (x_{i+1} - x_i) \le \sum_{n=1}^N \sum_{i=k_n}^{\ell_n - 1} (x_{i+1} - x_i) = \sum_{n=1}^N |J_n|.$$

Now $\sum_{n=1}^\infty |I_n| \ge \sum_{n=1}^N |I_n| = \sum_{n=1}^N \left(|J_n| - 2^{-(n-1)}\varepsilon\right) > b - a - 2\varepsilon$. This holds for every $\varepsilon > 0$, so $\sum_{n=1}^\infty |I_n| \ge b-a$. Hence $m^*[a,b] \ge b-a$. □

A subset E of $\mathbb{R}$ is said to be *null* if $m^*(E) = 0$.

Corollary 2.2. 1. *Any subset of a null set is null.*
2. *If E_n is a null set for $n = 1, 2, \dots$, then $\bigcup_{n=1}^\infty E_n$ is null.*
3. *Any countable subset of $\mathbb{R}$ is null.*

Proof. **[Direct proof of (2)]** Let $\varepsilon > 0$. There exist intervals $I_{r,n}$ such that $E_n \subseteq \bigcup_{r=1}^\infty I_{r,n}$ and $\sum_r |I_{r,n}| < \varepsilon 2^{-n}$. Now $\{I_{r,n} : r, n = 1, 2, \dots\}$ is a countable family of intervals covering $\bigcup E_n$, and $\sum_n \sum_r |I_{r,n}| < \sum_n \varepsilon 2^{-n} = \varepsilon$. Hence $m^*(\bigcup_n E_n) = 0$. □

Example 2.3. Let $C_0 = [0,1]$, $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$, $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, etc. In general, C_n is the union of 2^n disjoint closed intervals, each of length 3^{-n}, and C_{n+1} is obtained from C_n by deleting the open middle third of each of those intervals.

Let $C = \bigcap_{n=1}^\infty C_n$. Then C is a closed subset of $\mathbb{R}$, known as the *Cantor set*. Clearly, $m^*(C) \le 2^n 3^{-n}$ for each n. Letting $n \to \infty$ shows that C is null.

Let $x \in [0,1]$. Then $x \in C$ if and only if x has a ternary expansion $x = \sum_{n=1}^\infty a_n 3^{-n}$, where each $a_n = 0$ or 2. Then a variation of Cantor's proof shows that C is uncountable.

A property Q of real numbers is said to hold *almost everywhere* (a.e.) if the set of real numbers for which Q does not hold is a null set. For example, $\chi_C = 0$ a.e., i.e., $\chi_C(x) = 0$ for almost all x, because C is null.

Now let us consider the question whether m^* is countably additive.

Example 2.4. Let A be a subset of $[0, 1]$ with the following properties;

(i) $x, y \in A, x \neq y \implies x - y \notin \mathbb{Q}$;
(ii) For any $x \in [0, 1]$, there exists $q \in \mathbb{Q}$ such that $x + q \in A$.

Then

$$[0, 1] \subseteq \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (A - q) \subseteq [-1, 2].$$

Moreover, the sets $A - q$ are disjoint (as q varies), and there are countably many of them. If m^* is countably additive, then

$$1 = m^*[0, 1] \leq \sum_{q \in \mathbb{Q} \cap [-1,1]} m^*(A - q) = \sum_{q \in \mathbb{Q} \cap [-1,1]} m^*(A) \leq 3.$$

This is impossible.

Thus m^* is not countably additive, provided that such a set A exists. The additive group $\mathbb{R}$ is partitioned into the cosets of its additive subgroup $\mathbb{Q}$, and (i) and (ii) say that A contains exactly one member of each coset of $\mathbb{Q}$. The existence of such a set follows from the Axiom of Choice, an axiom of set theory beyond the basic axioms. This shows that it is impossible to prove that m^* is countably additive without using some weird axiom which contradicts the Axiom of Choice. On the other hand, it can be proved that it is impossible to show that m^* is not countably additive, using only the basic axioms of set theory.

This is bad news, but it is not so very bad because the badness occurs only with sets which cannot be explicitly described. So we can rescue things by restricting attention to a class of sets with good behaviour.

A subset E of $\mathbb{R}$ is said to be *(Lebesgue) measurable* if

$$m^*(A) = m^*(A \cap E) + m^*(A \setminus E)$$

for all subsets A of $\mathbb{R}$. Here, $A \setminus E = A \cap (\mathbb{R} \setminus E)$—it is not assumed that $E \subseteq A$.[1]

Let $\mathcal{M}_{\text{Leb}}$ be the set of all Lebesgue measurable subsets of $\mathbb{R}$.

Proposition 2.5. 1. *If E is null then $E \in \mathcal{M}_{\text{Leb}}$.*
2. *If I is any interval, then $I \in \mathcal{M}_{\text{Leb}}$.*
3. *If $E \in \mathcal{M}_{\text{Leb}}$, then $\mathbb{R} \setminus E \in \mathcal{M}_{\text{Leb}}$.*
4. *If $E_n \in \mathcal{M}_{\text{Leb}}$ for $n = 1, 2, \ldots$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{M}_{\text{Leb}}$.*

[1]The definition we use is the same definition as Capinski & Kopp and Zhongming Qian's lecture notes (2017) — this is known as the *Carthedory criterion* for measurability. Etheridge had a different definition, Stein & Shakarchi have another, Garling has another; and Priestley has yet another. All these definitions are equivalent, but this requires some work; we will see the equivalence of the definition above with that used by Stein and Shakarchi after Corollary 2.7 (but relying on your work proving the Lebsgue mesurable sets from a σ-algebra in Proposition 2.5).

5. *If* $E_n \in \mathcal{M}_{\text{Leb}}$ *for* $n = 1, 2, \ldots$ *and* $E_n \cap E_k = \emptyset$ *whenever* $n \neq k$*, then* $m^*\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} m^*(E_n)$.

The proofs are exercises (Q9, Sheet 1 for 1,2,4 and 5), or can be found in books such as Capinski & Kopp. (3) is almost trivial.

Note $\bigcap_{n=1}^{\infty} E_n = \mathbb{R} \setminus (\bigcup_{n=1}^{\infty} \mathbb{R} \setminus E_n)$, $\mathcal{M}_{\text{Leb}}$ is also closed under (finite or countable) intersections. The set A of Example 2.4 is not Lebesgue measurable.

Corollary 2.6. *All open subsets, and all closed subsets of* $\mathbb{R}$*, are Lebesgue measurable.*

Proof. Any open subset of $\mathbb{R}$ is a countable union of intervals (See the optional exercise: sheet 1 Q8). □

For $E \in \mathcal{M}_{\text{Leb}}$, we shall write $m(E)$ for $m^*(E)$. Then $m : \mathcal{M}_{\text{Leb}} \to [0, \infty]$ is countably additive.

The definition of Lebesgue measurability we have chosen to use is designed for use in the proof that the Lebesgue measurable sets are closed under countable unions. Also the Cartheodory condition generalises very nicely, and is an essential part of the Carthedory extension theorem which is a fundamental tool for producing measures (see B8.1: Probability Measure and Martingales, or Chapter 6.1 of Stein and Shakarchi). While the Carthedory condition is designed for use in proofs, it's hard to visualise being a condition quantified over all sets A. So we end with an alternative description of Lebesgue measurable sets:

Corollary 2.7. *Let* $E \subset \mathbb{R}$*. The following are equivalent.*

1. $E \in \mathcal{M}_{\text{Leb}}$;
2. *for all* $\varepsilon > 0$*, there exists an open set* $U \supseteq E$ *with* $m^*(U \setminus E) < \varepsilon$;
3. *there exist open sets* $(U_n)_{n=1}^{\infty}$ *with* $E \subset \bigcap_{n=1}^{\infty} U_n$ *and* $m^*(\bigcap_{n=1}^{\infty} U_n \setminus E) = 0$.

Proof. $1 \Rightarrow 2$: Suppose first that $m(E) < \infty$. Then we can find countably many open intervals $(I_n)_{n=1}^{\infty}$ with $E \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_n |I_N| < m(E) + \varepsilon$. Let $U = \bigcup_{n=1}^{\infty} I_n$ so $E \subseteq U$ and $m(E) \leq m(U) \leq \sum_n |I_n| < m(E) + \varepsilon$. Applying the definition of measurability with $A = U$,[2] we get $m(U) = m(E) + m(U \setminus E) < m(E) + \varepsilon$, so as $m(E) < \infty$, $m(U \setminus E) < \varepsilon$.

When $m(E) = \infty$, let $E_n = E \cap [-n, n]$ which is measurable with $m(E_n) < \infty$ and $E = \bigcup_{n=1}^{\infty} E_n$. For $\varepsilon > 0$ use the previous paragraph to find open sets U_n with $U_n \supseteq E_n$ and $m(U_n \setminus E_n \setminus) < 2^{-n}\varepsilon$. For $U = \bigcup_{n=1}^{\infty} U_n$ we have $E \subseteq U$ and $m(U \setminus E) \leq m(\bigcup_{n=1}^{\infty}(U_n \setminus E_n)) \leq \sum_{n=1}^{\infty} m(U_n \setminus E_n) < \varepsilon$.

$2 \Rightarrow 3$: For each n, use 2 to find U_n open with $E \subset U_n$ and $m^*(U_n \setminus E) < 1/n$. Then $m^*(\bigcup_{n=1}^{\infty} U_n \setminus E) = 0$.

$3 \Rightarrow 1$. Given such U_n it follows that $\bigcap_{n=1}^{\infty} U_n$ is Lebesgue measurable (from Proposition 2.5 and Corollary 2.6), and $\bigcap_{n=1}^{\infty} U_n \setminus E$ is null so Lebesgue measurable (by Proposition 2.5). Then $E = \bigcap_{n=1}^{\infty} U_n \setminus (\bigcap_{n=1}^{\infty} U_n \setminus E)$ is Lebesgue measurable (by Proposition 2.5). □

[2]Or as U and E are both measurable, using finite additivity

Condition 2 of Corollary 2.7 is the definition of measurability used by Stein and Shakarchi.

Finally, let's end this section by discussing the case of $\mathbb{R}^n$ for $n \geq 2$ which we will use in Fubini's theorem later in the course. We define a *rectangle* R in $\mathbb{R}^n$ to be a product of intervals $I_1 \times I_2 \times \cdots \times I_n$ and let $|R| = \prod_{i=1}^n |I_n|$. Then for $A \subseteq \mathbb{R}^n$ the Lebesgue outer measure of A is defined by $m^*(A) = \inf\{\sum_{s=1}^\infty m^*(R_s) : A \subseteq \bigcup_{s=1}^\infty R_s,\ R_s \text{ rectangles}\}$ and proceed as in $\mathbb{R}$ to define the Lebesgue measurable sets via the Catheodory condition. Note that, for example, for any $E \subseteq \mathbb{R}$, $E \times \{0\}$ is null in $\mathbb{R}^2$. While the process is the same, there's a couple of small details which are a little trickier in $\mathbb{R}^n$ for $n \geq 2$: it is a bit more fiddly to formalise the (geometrically clear) fact that $|R| \leq \sum_{i=1}^K |R_i|$ whenever a bounded rectangle R has $R \subseteq \bigcup_{i=1}^K R_i$ for bounded rectangles R_i in $\mathbb{R}^n$ than the corresponding result in $\mathbb{R}$.[3]. Also while open subsets in $\mathbb{R}$ are countable disjoint unions of open intervals (see exercises), open subsets of $\mathbb{R}^n$ are not countable disjoint unions of open rectangles (think about what happens on the boundary); you have to allow for unions of rectangles which only intersect in the boundaries. Nevertheless Corollary 2.7 and the characterisations of the next paragraph hold equally well in $\mathbb{R}^n$.

Specialising to $\mathbb{R}^2$ for notational purposes, one can then show that if $E_1, E_2 \subseteq \mathbb{R}$ are measurable, then so is $E_1 \times E_2$: one finds $G_1 \supseteq E_1$ and $G_2 \supseteq E_2$ with G_i a countable intersection of open sets, and $G_i \setminus E_i$ is null. Then $G_1 \times G_2$ is a countable intersection of open sets and you can check that $(G_1 \times G_2) \setminus (E_1 \times E_2)$ is null in $\mathbb{R}^2$ (see Stein and Shakarchi Proposition 3.3.6).[4]

3. Measure spaces and measurable functions

Let Ω be any set, and $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. We say that $\mathcal{F}$ is a *σ-algebra* (or *σ-field*) on Ω if:

(i) $\emptyset \in \mathcal{F}$,
(ii) If $E \in \mathcal{F}$, then $\Omega \setminus E \in \mathcal{F}$,
(iii) If $E_n \in \mathcal{F}$ for $n = 1, 2, \dots$, then $\bigcup_{n=1}^\infty E_n \in \mathcal{F}$.

Then $(\Omega, \mathcal{F})$ is a *measurable space*, and sets in $\mathcal{F}$ are *$\mathcal{F}$-measurable*. As before, $\bigcap E_n \in \mathcal{F}$ if $E_n \in \mathcal{F}$ for $n = 1, 2, \dots$.

A *measure* on $(\Omega, \mathcal{F})$ is a function $\mu : \mathcal{F} \to [0, \infty]$ such that

(i) $\mu(\emptyset) = 0$,
(ii) $\mu(\bigcup_{n=1}^\infty E_n) = \sum_{n=1}^\infty \mu(E_n)$ whenever E_n are disjoint sets in $\mathcal{F}$.

Then $(\Omega, \mathcal{F}, \mu)$ is a *measure space.*

A measure μ is *finite* if $\mu(\Omega) < \infty$; μ is a *probability measure* if $\mu(\Omega) = 1$.

Examples 3.1. 1. $(\mathbb{R}, \mathcal{M}_{\text{Leb}}, m)$ is a measure space. Also, $([0,1], \mathcal{M}_{\text{Leb}}|_{[0,1]}, m)$ is a probability space, where $\mathcal{M}_{\text{Leb}}|_{[0,1]}$ is the set of all Lebesgue measurable subsets of $[0, 1]$.

---

[3]Though in the proof that $m^*(R) = |R|$ for closed and bounded rectangles, compactness is still the key trick.

[4]The converse is false: consider a non-measurable subset $A \subseteq [0, 1]$, then $A \times \{0\}$ is null in $\mathbb{R}^2$ so measurable.

2. Let Ω be any set, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mu(E) = |E|$ (the number of elements of E). This is a measure space; μ is *counting measure* on Ω.
3. In probability theory, let Ω be a sample space of all possible outcomes, $\mathcal{F}$ be the collection of all events E, and $\mathbb{P}(E)$ be the probability that event E occurs. Then $\mathbb{P}$ is a probability measure on $(\Omega, \mathcal{F})$, and $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space.
4. Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing function. Note that F may be discontinuous, but its left and right limits exist at each point. We assume that $F(x) = \lim_{y\to x+} F(y)$ for all x (without essential loss). Define
$$m_F(a,b] = F(b) - F(a),$$
$$m_F^*(E) = \inf\left\{\sum_{n=1}^{\infty} m_F(J_n) : J_n = (a_n, b_n], E \subseteq \bigcup_{n=1}^{\infty} J_n\right\}.$$
Then m_F^* has similar properties to m^*, but one has to be aware that $m_F^*(a,b) = F(b-) - F(a)$, $m_F^*([a,b]) = F(b) - F(a-)$; and $m_F^*(\{x\}) = 0$ if and only if F is continuous at x. One can then define a σ-algebra $\mathcal{M}_F$, containing all intervals, in the same way as $\mathcal{M}_{\text{Leb}}$, and m_F^* is a measure, written m_F on $\mathcal{M}_F$. This is the *Lebesgue-Stieltjes measure associated with F.*

Proposition 3.2. *Let $(\Omega, \mathcal{F}, \mu)$ be a measure space.*

1. *If $A, B \in \mathcal{F}$ and $A \subseteq B$, then $\mu(A) \le \mu(B)$.*
2. *If $A_n \in \mathcal{F}$ and $A_n \subseteq A_{n+1}$ for all n, then $\mu(\bigcup_n A_n) = \lim_{n\to\infty} \mu(A_n)$.*
3. *If $A_n \in \mathcal{F}$ and $A_n \supseteq A_{n+1}$ for all n and $\mu(A_1) < \infty$, then $\mu(\bigcap_n A_n) = \lim_{n\to\infty} \mu(A_n)$.*

Proof. (1) Since $B = A \cup (B \setminus A)$ (disjoint union), $\mu(B) = \mu(A) + \mu(B \setminus A) \ge \mu(A)$.

(2) Let $A_1' = A_1$ and $A_r' = A_r \setminus A_{r-1}$ for $r \ge 2$. Then $A_n = \bigcup_{r=1}^n A_r'$, $\bigcup_{n=1}^\infty A_n = \bigcup_{r=1}^\infty A_r'$ (disjoint unions), so
$$\mu(\bigcup A_n) = \sum_{r=1}^{\infty} \mu(A_r') = \lim_{n\to\infty} \sum_{r=1}^{n} \mu(A_r') = \lim_{n\to\infty} \mu(A_n).$$
(3) is an exercise. □

We can produce σ-algebras starting with any collection of sets.

Proposition 3.3. *Let Ω be a set, and $\mathcal{B} \subseteq \mathcal{P}(\Omega)$. Then there is a unique σ-algebra $\mathcal{F}_{\mathcal{B}}$ on Ω satisfying:*

(i) $\mathcal{F}_{\mathcal{B}}$ is a σ-algebra and $\mathcal{B} \subseteq \mathcal{F}_{\mathcal{B}}$,
(ii) If $\mathcal{F}$ is σ-algebra on Ω and $\mathcal{B} \subseteq \mathcal{F}$ then $\mathcal{F}_{\mathcal{B}} \subseteq \mathcal{F}$.

Proof. We let $\mathcal{F}_{\mathcal{B}}$ be the intersection of all σ-algebras on Ω which contain $\mathcal{B}$ (which you should check is a σ-algebra, so that (i) holds). By definition (ii) holds. Notice that if $\mathcal{F}_{\mathcal{B}}'$ is another such σ-algebra, then applying (ii) for $\mathcal{F}_{\mathcal{B}}$ we have $\mathcal{F}_{\mathcal{B}} \subseteq \mathcal{F}_{\mathcal{B}}'$. But reversing the roles, we can apply (ii) for $\mathcal{F}_{\mathcal{B}}'$ giving $\mathcal{F}_{\mathcal{B}}' \subseteq \mathcal{F}_{\mathcal{B}}$. □

The σ-algebra $\mathcal{M}_{\text{Bor}}$ generated by the intervals is the *Borel σ-algebra* on $\mathbb{R}$. It can be described as the class of all subsets of $\mathbb{R}$ which can be obtained from intervals in a countable number of steps, each of which is one of taking the complement of a set, taking a countable union of sets, or a countable intersection of sets. However this has to be treated with caution, because it is not necessarily possible to obtain a given Borel set by performing the countable number of steps in a single sequence.

Proposition 3.4. 1. *Let $\mathcal{B}$ be any one of the following classes of subsets of $\mathbb{R}$.*

- *(i) All intervals*
- *(ii) All intervals of the form (a, ∞)*
- *(iii) All intervals of the form $[a, b]$*
- *(iv) All open sets.*

Then $\mathcal{M}_{\text{Bor}}$ is the smallest σ-algebra on $\mathbb{R}$ containing $\mathcal{B}$.

2. $\mathcal{M}_{\text{Bor}} \neq \mathcal{M}_{\text{Leb}}$.
3. *If $E \in \mathcal{M}_{\text{Leb}}$ there exist $A, B \in \mathcal{M}_{\text{Bor}}$ such that $A \subseteq E \subseteq B$ and $B \setminus A$ is null (so $E \setminus A$ and $B \setminus E$ are null).*

Proof. (1) is an exercise involving showing each interval can be obtained from members of $\mathcal{B}$, and each member of $\mathcal{B}$ can be obtained from intervals (see Sheet 2, Q3 for (ii)). (2) is deeper, and is discussed in the appendix to Capinski & Kopp. (3) is Theorem 2.28 in Capinski & Kopp and can be deduced from Corollary 2.7. □

Let $(\Omega, \mathcal{F})$ be a measurable space. A function $f : \Omega \to \mathbb{R}$ is *$\mathcal{F}$-measurable* if $f^{-1}(I) \in \mathcal{F}$ for each interval I.

Proposition 3.5. *Let $\mathcal{B}$ be any one of the classes of subsets of $\mathbb{R}$ listed in Proposition 3.4. Let $f : \Omega \to \mathbb{R}$. Then f is $\mathcal{F}$-measurable if and only if $f^{-1}(G) \in \mathcal{F}$ for all $G \in \mathcal{M}_{\text{Bor}}$ or for all $G \in \mathcal{B}$.*

Proof. It is easily verified that $f_*(\mathcal{F}) := \{G \subseteq \mathbb{R} : f^{-1}(G) \in \mathcal{F}\}$ is a σ-algebra on $\mathbb{R}$. Hence if $\mathcal{B}$ generates the σ-algebra $\mathcal{M}_{\text{Bor}}$, then the result holds. □

In this course, we shall usually take $(\Omega, \mathcal{F})$ to be $(\mathbb{R}, \mathcal{M}_{\text{Leb}})$ or minor variants, but much of this section will apply to the general case as well. We may refer to $\mathcal{M}_{\text{Leb}}$-measurable functions simply as measurable functions, for simplicity; or as Lebesgue measurable functions. We shall also be interested in cases where Ω is an interval (or a Lebesgue measurable subset) and $\mathcal{F} = \mathcal{M}_{\text{Leb}}|_\Omega = \{E \in \mathcal{M}_{\text{Leb}} : E \subseteq \Omega\}$. However, $f : \Omega \to \mathbb{R}$ is $\mathcal{M}_{\text{Leb}}|_\Omega$-measurable if and only if $\tilde{f} : \mathbb{R} \to \mathbb{R}$ is measurable, where $\tilde{f}(x) = f(x)$ for $x \in \Omega$, and $f(x) = 0$ otherwise. So we may state results just for functions defined on $\mathbb{R}$.

Recall from the Analysis courses that $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if $f^{-1}(G)$ is open for every open set (or open interval) G. By Proposition 3.5, we have that f is (Lebesgue) measurable if and only if $f^{-1}(G)$ is (Lebesgue) measurable for every open set (or open interval) G.

Examples 3.6. 1. Constant functions are measurable.

2. The characteristic function χ_A of a subset A of $\mathbb{R}$ is measurable if and only if A is a measurable set. In particular, if A is as in Example 2.4, then χ_A is not (Lebesgue) measurable.
3. Continuous functions $f : \mathbb{R} \to \mathbb{R}$ are measurable.
4. Monotone functions $f : \mathbb{R} \to \mathbb{R}$ are measurable.

5. If f is continuous a.e., then f is measurable.
6. If $f : \mathbb{R} \to \mathbb{R}$ is (Lebesgue) measurable and $g = f$ a.e., then g is (Lebesgue) measurable.
7. In probability theory, measurable functions are called random variables.

It follows from the definition of measurable functions and Example 3.6(2) that the existence of a non-measurable function is equivalent to the existence of a non-measurable set. So their existence depends on the Axiom of Choice. Thus, we have the following:

Fact of Life. ALL FUNCTIONS $f : \mathbb{R} \to \mathbb{R}$ THAT CAN BE EXPLICITLY DEFINED ARE LEBESGUE MEASURABLE.

This is not exactly a mathematical theorem—it becomes one if one interprets "explicitly defined" in the right technical way. It is a true statement about the real world: a non-measurable function involves some non-explicit choice process. Priestley compares the existence of non-measurable functions to the existence of yetis.

Nevertheless, measurability is a real issue in some more advanced mathematics, because:

(a) One may be interested not in Lebesgue measurability of functions f on $\mathbb{R}$, but in measurability on some other measurable space $(\Omega, \mathcal{F})$. This occurs frequently in time-dependent probability theory, where $\mathcal{F}_t$ is the class of all events depending only on past history up to time t, not on the future (cf. Part B courses on martingales and stochastic calculus).

(b) One may be interested in functions f which are not real-valued, but take values in an infinite-dimensional space. Then measurability is a real issue in many areas of analysis, although you probably won't see this in your undergraduate course.

So it is useful to accumulate general results about measurable functions, even if we only state them for functions $f : (\mathbb{R}, \mathcal{M}_{\text{Leb}}) \to \mathbb{R}$.

Proposition 3.7. *Let f and g be measurable functions from $\mathbb{R}$ to $\mathbb{R}$. The following functions are measurable:*

$$f + g, fg, \max(f, g), \quad h \circ f \text{ for any continuous function } h.$$

For example, αf is measurable, where $\alpha \in \mathbb{R}$.

Proof. For example,

$$(f + g)^{-1}(a, \infty) = \bigcup_{q \in \mathbb{Q}} f^{-1}(q, \infty) \cap g^{-1}(a - q, \infty).$$

If G is open in $\mathbb{R}$, then $h^{-1}(G)$ is open. Since f is measurable, $f^{-1}(h^{-1}(G))$ is measurable, i.e., $(h \circ f)^{-1}(G)$ is measurable. ☐

In fact, it suffices in Proposition 3.7 that h should be Borel measurable.

Now we want to consider limits and suprema of sequences of functions (f_n). Even if each f_n is real-valued, the resulting functions may take the values ∞ and $-\infty$.

A function $f : \mathbb{R} \to [-\infty, \infty]$ is *measurable* if $f^{-1}(a, \infty] \in \mathcal{M}_{\text{Leb}}$ for all $a \in \mathbb{R}$; equivalently $f^{-1}(B) \in \mathcal{M}_{\text{Leb}}$ for all $B \in \mathcal{M}_{\text{Bor}}$ and $f^{-1}(\{\infty\}) \in \mathcal{M}_{\text{Leb}}$; equivalently, $\arctan \circ f$ is measurable, where $\arctan : [-\infty, \infty] \to [-\pi/2, \pi/2]$ is the inverse tan function.

Proposition 3.8. *Let (f_n) be a sequence of measurable functions from $\mathbb{R} \to [-\infty, \infty]$. Then the following functions are measurable:*

$$\sup_n f_n, \inf_n f_n, \limsup_{n\to\infty} f_n, \liminf_{n\to\infty} f_n.$$

Hence, if $f(x) = \lim_{n\to\infty} f_n(x)$ a.e., then f is measurable.

Proof. First,

$$(\sup f_n)^{-1}(a, \infty] = \bigcup_n f_n^{-1}(a, \infty] \in \mathcal{M}_{\text{Leb}}.$$

Then

$$\begin{aligned} \inf f_n &= -\sup(-f_n), \\ \limsup f_n &= \inf g_m, \text{ where } g_m = \sup_{n \geq m} f_n. \end{aligned}$$

□

A function $\phi : \mathbb{R} \to \mathbb{R}$ is *simple* if it is measurable and it takes only finitely many real values. So χ_E is simple if $E \in \mathcal{M}_{\text{Leb}}$. If ϕ, ψ are simple, then so are $\phi + \psi$, $\phi.\psi$, $\alpha\phi$, $\max(\phi, \psi)$, $h \circ \phi$ for any function h.

Any function of the form $\sum_{j=1}^n \beta_j \chi_{E_j}$, where $\beta_j \in \mathbb{R}$ and $E_j \in \mathcal{M}_{\text{Leb}}$ is simple. On the other hand, if ϕ is simple with non-zero values $\alpha_1, \dots, \alpha_k$, and $B_i = \phi^{-1}(\{\alpha_i\})$, then B_i is measurable, and

$$\phi = \sum_{i=1}^k \alpha_i \chi_{B_i}. \qquad (*)$$

In this form, we have

(i) α_i are distinct and non-zero,
(ii) B_i are disjoint.

If these additional properties hold, then (*) is unique (up to reordering of the terms). We shall then say that ϕ is in *standard*, or *canonical*, form. For example, the standard form of $\chi_{(0,2)} + \chi_{[1,3]}$ is $1\chi_{(0,1)\cup[2,3]} + 2\chi_{[1,2)}$.

In defining simple functions, some authors insist that the sets B_i, corresponding to non-zero α_i, must be bounded [Etheridge] or of finite measure [Stein & Shakarchi]. [Garling and Priestley avoid introducing simple functions.]

Examples 3.9. 1. Any step function is a simple function—for a step function, the sets B_i in the standard representation must be finite unions of bounded intervals (or single points).

2. The function $\chi_{\mathbb{Q}\cap[0,1]}$ is a simple function but it is not a step function.

Proposition 3.10. *Let $f : \mathbb{R} \to [0, \infty]$ be measurable. There is an increasing sequence (ϕ_n) of non-negative simple functions ϕ_n such that*

$$f(x) = \lim_{n\to\infty} \phi_n(x)$$

for all $x \in \mathbb{R}$.

Proof. For $n = 1, 2, \ldots$ and $k = 0, 1, 2, \ldots, 4^n - 1$, let

$$B_{k,n} = \left\{x : k2^{-n} \le f(x) < (k+1)2^{-n}\right\}.$$

Let

$$\phi_n(x) = \begin{cases} k2^{-n} & \text{if } x \in B_{k,n} \text{ for some (unique) } k, \\ 2^n & \text{if } f(x) \ge 2^n. \end{cases}$$

Then $\phi_n \le \phi_{n+1}$, $\phi_n \le f$, $\phi_n(x) > f(x) - 2^{-n}$ for all sufficiently large n if $f(x) < \infty$, and $\phi_n(x) = 2^n$ for all n if $f(x) = \infty$. □

Notice here that the approximating simple functions are constructed by taking horizontal strips, unlike Prelims where vertical strips were used.

Theorem 3.11. *A function $f : \mathbb{R} \to \mathbb{R}$ is measurable if and only if there is a sequence of step functions ψ_n such that $f = \lim \psi_n$ a.e.*

Proof. Stein & Shakarchi, Theorem 4.3, p.32. □

4. The Lebesgue integral: non-negative functions

We now start to define our notion of the integral. In contrast to Riemann's theory which simultaneously considers upper and lower approximations to the area under the curve, in Lebesgue's theory we approximate area that lies above the x-axis from below, and area below the x-axis from above. This leads us to split any function into its positive and negative parts, and integrate these separately. In this section, we'll develop the theory of integration for non-negative functions, and turn to the general case in the following section.

For a non-negative simple function ϕ with standard form $\sum_{i=1}^k \alpha_i \chi_{B_i}$ (so $\alpha_i > 0$), the *integral* of ϕ is defined to be:

$$\int_{\mathbb{R}} \phi = \int_{-\infty}^{\infty} \phi(x)\, dx = \sum_{i=1}^k \alpha_i m(B_i).$$

Note that $\int \phi < \infty$ if and only if $m(B_i) < \infty$ for each i.

Proposition 4.1. *Let ϕ, ψ be non-negative simple functions, $\alpha \in [0, \infty)$.*

1. *If $\phi = \sum_{j=1}^n \beta_j \chi_{E_j}$ where $\beta_j \ge 0$ and E_j are measurable (but not necessarily in standard form), then $\int \phi = \sum_j \beta_j m(E_j)$.*
2. $\int (\phi + \psi) = \int \phi + \int \psi$, $\quad \int \alpha\phi = \alpha \int \phi$.
3. *If $\phi \le \psi$ then $\int \phi \le \int \psi$.*

The first statement of Proposition 4.1 is not completely obvious, but fortunately it is true! [Capinski & Kopp define $\int \phi$ to be $\sum_j \beta_j m(E_j)$, ignoring the question whether this is well-defined.]

For a non-negative measurable function $f : \mathbb{R} \to [0, \infty]$, we define the *integral* of f to be

$$\int_{\mathbb{R}} f = \sup \left\{ \int_{\mathbb{R}} \phi : \phi \text{ simple, } 0 \leq \phi \leq f \right\}.$$

For a measurable subset E of $\mathbb{R}$, we define

$$\int_E f = \int_{\mathbb{R}} f \chi_E.$$

For a measurable function $f : E \to [0, \infty)$, we define $\int_E f = \int_{\mathbb{R}} \tilde{f}$, where $\tilde{f}$ agrees with f on E and is 0 on $\mathbb{R} \setminus E$.

Notice that for non-negative measurable functions $f : E \to \mathbb{R}$ we allow $\int_E f$ to take the value ∞, so the integral is always defined, but it may not be finite.[5] We say that f is *integrable over* E if $\int_E f < \infty$.

This definition of integral corresponds to the lower integral in Prelims, but with simple functions replacing step functions. If the Monotone Convergence Theorem is to be true, then Proposition 3.10 shows that the integral must equal this supremum, but it is still necessary to show that our definition has good properties.

It is clear from the definition of integral that

(i) $\int \alpha f = \alpha \int f$ $(\alpha \geq 0)$;
(ii) If $f \leq g$, then $\int f \leq \int g$,

The first things to establish are

(iii) $\int (f + g) = \int f + \int g$,
(iv) The Monotone Convergence Theorem.

Theorem 4.2. [Monotone Convergence Theorem] *If (f_n) is an increasing sequence of non-negative measurable functions and $f = \lim_{n\to\infty} f_n$, then $\int f = \lim_{n\to\infty} \int f_n$.*

This is the first of our three big convergence theorems. We'll give a slight strengthening of the theorem as Theorem 6.1.

Proof. Since $f_n \leq f$, it is immediate that $\sup_n \int f_n \leq \int f$.

For the reverse inequality, we consider a simple function ϕ such that $0 \leq \phi \leq f$. We have to show that $\int \phi \leq \lim_{n\to\infty} \int f_n$. It then follows from the definition of $\int f$ that $\int f \leq \lim_{n\to\infty} \int f_n$.

Take $\alpha \in (0, 1)$, and let

$$B_n = \{x : f_n(x) \geq \alpha \phi(x)\}.$$

[5] You should compare this with the next section: for functions taking both positive and negative values we need the function to be integrable before we define $\int f$.

Then B_n is measurable (since $f_n - \alpha\phi$ is measurable), $B_n \subseteq B_{n+1}$ and $\bigcup_{n=1}^\infty B_n = \mathbb{R}$ (for each x, either $\phi(x) = 0$ or $f(x) > \alpha\phi(x)$). Since $\alpha\phi\chi_{B_n} \le f_n\chi_{B_n} \le f_n$,

$$(*) \qquad \alpha \int_{B_n} \phi \le \int_{\mathbb{R}} f_n.$$

If $\phi = \sum_{i=1}^k \beta_i \chi_{E_i}$, then

$$\int_{B_n} \phi = \sum_{i=1}^k \beta_i m(E_i \cap B_n) \to \sum_{i=1}^k \beta_i m(E_i) = \int_{\mathbb{R}} \phi$$

as $n \to \infty$, by Proposition 3.2(2). Taking limits in (*),

$$\alpha \int_{\mathbb{R}} \phi \le \lim_{n\to\infty} \int_{\mathbb{R}} f_n.$$

Letting $\alpha \to 1-$ gives the required inequality. □

We have not specified the range of integration. It could be $\mathbb{R}$, or it could be a fixed interval I. We can also apply the MCT when the the range of integration depends on n, by taking f_n to be 0 elsewhere. See Example 4.8.

Corollary 4.3. **[Baby MCT]** *Let f be a non-negative measurable function, (E_n) be an increasing sequence of measurable sets, and $E = \bigcup_{n=1}^\infty E_n$. Then $\int_E f = \sup_n \int_{E_n} f = \lim_{n\to\infty} \int_{E_n} f$, and so f is integrable over E if $\sup_n \int_{E_n} f < \infty$.*

Proof. Apply Theorem 4.2 with $f_n = f\chi_{E_n}$, noting that $\chi_{E_n} \le \chi_{E_{n+1}}$ and $f \ge 0$, so $f_n \le f_{n+1}$ and $\chi_E(x) = \lim_{n\to\infty} \chi_{E_n}(x)$. □

The baby version of the MCT will be used a lot in order to use the fundamental theorem of calculus to compute integrals on closed and bounded sets using the theory developed in prelims.

Corollary 4.4. *For non-negative measurable functions f and g,*

$$\int (f+g) = \int f + \int g.$$

Proof. Let (ϕ_n) and ψ_n be increasing sequences of non-negative simple functions, converging pointwise to f and g respectively (Proposition 3.10). Then $(\phi_n + \psi_n)$ is an increasing sequence, converging to $f + g$. By MCT and Proposition 4.1(2),

$$\int (f+g) = \lim_{n\to\infty} \int (\phi_n+\psi_n) = \lim_{n\to\infty} \left(\int \phi_n + \int \psi_n \right) = \lim_{n\to\infty} \int \phi_n + \lim_{n\to\infty} \int \psi_n = \int f + \int g.$$

□

Corollary 4.5. **[MCT for Series]** *Let f_n be non-negative measurable functions and $f = \sum_{n=1}^\infty f_n$. Then $\int f = \sum_{n=1}^\infty \int f_n$. In particular, f is integrable if and only if $\sum_n \int f_n < \infty$.*

Proof. Let $g_n = \sum_{r=1}^n f_r$, and apply MCT. □

In order to give any interesting examples, we need to show that the integrals just defined agree with the Riemann integral initially for continuous functions on closed bounded intervals. We will come back to this in Section 5, but record the result here for continuous functions for use in the next few examples.[6]

Corollary 4.6. *Let $f : [a, b] \to [0, \infty)$ be continuous. Then the Lebesgue integral $\int_{[a,b]}^{\mathcal{L}} f$ as defined above equals the Riemann integral $\int_{[a,b]}^{\mathcal{R}} f$ as defined in first-year Integration.*

Example 4.7. Consider $f(x) = (1 - x)^{-1/2}$ on $(0, 1)$. By Baby MCT (Corollary 4.3), Corollary 4.6 and FTC (from Prelims),

$$\int_0^1 (1-x)^{-1/2}\,dx = \lim_{n\to\infty} \int_0^{1-\frac{1}{n}} (1-x)^{-1/2}\,dx = \lim_{n\to\infty} 2(1 - n^{-1/2}) = 2.$$

For $0 \le x < 1$, the Binomial Theorem with exponent $-1/2$ or Taylor's Theorem in complex analysis gives

$$(1-x)^{-1/2} = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} x^n.$$

By Corollary 4.5 and FTC,

$$\int_0^1 (1-x)^{-1/2}\,dx = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n (n!)^2} \int_0^1 x^n\,dx = \sum_{n=0}^{\infty} \frac{(2n)!}{4^n n!(n+1)!}\,.$$

The fact that the series above converges to 2 can be obtained directly from the Binomial Expansion of $(1 - x)^{1/2}$, via Abel's continuity theorem (A2 lecture notes MT 2019, Theorem 13.24).

Example 4.8. Consider $\displaystyle\int_0^{n\pi} \left(\cos \frac{x}{2n}\right) x^2 e^{-x^3}\,dx$. It is not obvious how to evaluate the integral for a given value of n, but we can use the MCT to find the limit of the integrals, as $n \to \infty$, as follows.

Let

$$f_n(x) = \left(\cos \frac{x}{2n}\right) x^2 e^{-x^3} \chi_{[0,n\pi]}(x) = \begin{cases} \left(\cos \frac{x}{2n}\right) x^2 e^{-x^3} & \text{if } 0 \le x \le n\pi \\ 0 & \text{otherwise.} \end{cases}$$

Fix n for a moment. We wish to show that $f_n(x) \le f_{n+1}(x)$ for all x. If $0 \le x \le n\pi$, then $0 \le \cos \dfrac{x}{2n} \le \cos \dfrac{x}{2(n+1)}$, so $f_n(x) \le f_{n+1}(x)$. If $n\pi < x \le (n+1)\pi$, then $f_n(x) = 0 \le f_{n+1}(x)$. If $x > (n+1)\pi$ (or if $x < 0$), then $f_n(x) = 0 = f_{n+1}(x)$. Thus we have established our claim that $f_n(x) \le f_{n+1}(x)$ for all x.

---

[6] This is easier than the result in section 5, as continuous functions are automatically measurable. So one can prove this by choosing a sequence of partitions given by repeatedly bisecting $[a, b]$, and taking the step functions associated to the lower Riemann sums for these partitions, we obtain an increasing sequence (ϕ_n) of step functions such that $\lim_{n\to\infty} \phi_n(x) = f(x)$ for all $x \in [a, b]$ (continuity ensures that one has convergence everywhere) and $\lim_{n\to\infty} \int_a^b \phi_n = \int_{[a,b]}^{\mathcal{R}} f$. By the MCT (Theorem 4.2), $\lim_{n\to\infty} \int_a^b \phi_n = \int_{[a,b]}^{\mathcal{L}} f$. As you'll see in Section 5, for a general Riemann integrable f, we can arrange for the ϕ_n to converge to f almost everywhere, and then the same result holds.

Noting that $f_n(x) \to f(x) = x^2 e^{-x^3}$ for all $x \geq 0$, the MCT gives

$$\lim_{n\to\infty} \int_0^{n\pi} \left(\cos \frac{x}{2n}\right) x^2 e^{-x^3}\, dx = \lim_{n\to\infty} \int_0^n f_n(x)\, dx = \int_0^\infty x^2 e^{-x^3}.$$

This can be computed using the Baby MCT at the first step below, and the FTC at the second:

$$\int_0^\infty x^2 e^{-x^3}\, dx = \lim_{n\to\infty} \int_0^n x^2 e^{-x^3}\, dx = \lim_{n\to\infty} \frac{1 - e^{-n^3}}{3} = \frac{1}{3}.$$

5. The Lebesgue integral: general functions

Now we turn to integrability of functions which are not necessarily non-negative.

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

$$f^+ = \max(f, 0), \qquad f^- = \max(-f, 0).$$

Note that f^+ and f^- are measurable and non-negative, and

$$f = f^+ - f^-, \qquad |f| = f^+ + f^-.$$

We say that f is *integrable* if f is measurable and $\int f^+$ and $\int f^-$ are **both finite**. Notice that this requirement prevents any problems with $\infty - \infty$.[7] Then the *integral* of f is

$$\int f = \int f^+ - \int f^-.$$

Moreover, f is *integrable over a measurable subset* E if $f\chi_E$ is integrable. If $f : E \to [-\infty, \infty]$, then f is *integrable over* E if $\tilde{f}$ is integrable over $\mathbb{R}$. We write $f \in \mathcal{L}^1(E)$ to mean that f is integrable over E.

Proposition 5.1. 1. *If f is integrable, then $|f|$ is integrable.*
2. *If f is measurable and $|f|$ is integrable, then f is integrable.*
3. *(Comparison Test) If f is measurable and $|f| \leq g$ for some integrable function g, then f is integrable. If $|f| \geq g \geq 0$ for some measurable function g which is not integrable, then f is not integrable.*
4. *If f, g are both integrable and $f + g$ is defined, then $f + g$ is integrable and $\int(f + g) = \int f + \int g$. For $\alpha \in \mathbb{R}$, αf is integrable and $\int \alpha f = \alpha \int f$. If $f \leq g$, then $\int f \leq \int g$.*
5. *If f is integrable and $g = f$ a.e., then g is integrable and $\int g = \int f$.*
6. *If f is integrable then $f(x) \in \mathbb{R}$ a.e.*
7. *If f is integrable and $\int |f| = 0$ then $f(x) = 0$ a.e.*
8. *If f is integrable over a measurable set E and (E_n) is an increasing sequence of measurable sets with $\bigcup_{n=1}^\infty E_n = E$ then $\int_E f = \lim_{n\to\infty} \int_{E_n} f$.*

Proof. (1) and (2) follow from $\int f^\pm \leq \int |f| = \int f^+ + \int f^-$. (3) follows from $|f| \leq g \implies \int |f| \leq \int g$. (4) follows from $(f + g)^\pm \leq f^\pm + g^\pm$ and $(f + g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$. (5): Since $|g - f| = 0$ a.e., any simple function ϕ with $0 \leq \phi \leq |g - f|$ is a.e. 0, so its integral is 0. Hence $\int |g - f| = 0$. (6), (7): Exercise, Sheet 2 Q9. (8): Apply Baby MCT to f^+ and f^-. □

[7] It is possible to make sense of the quantity $\int f$ if one only has that one of $\int f^+$ or $\int f^-$ is finite, but this notion would not be well behaved — for example we definitely want the sum of two integrable functions to be integrable.

By (5), changing a function on a null set does not affect integrability. So if we have a function defined a.e., we can talk about it being integrable by considering any extension of f—for example, the extension by 0. Also, integrability over $[a,b]$ is the same as integrability over (a,b).

The following are corollaries of the Comparison Test.

Corollary 5.2. 1. *If g is integrable and h is bounded and measurable, then hg is integrable.*
2. *If g is integrable over $\mathbb{R}$, then g is integrable over any measurable subset of $\mathbb{R}$.*
3. *If h is a bounded measurable function, then h is integrable over any measurable subset of finite measure.*

Proof. These follow from the Comparison Test, using

$$|g.h| \le c|g|, \quad |g\chi_E| \le |g|, \quad |h\chi_E| \le c\chi_E. \quad \square$$

Apart from Corollary 4.6, almost all the theory in Section 4 up to this point applies to general measure spaces. Now we make some comments which are specific to the case of Lebesgue measure.

Firstly, as promised in Section 4, the Lebesgue integral is more general than the Riemann (Prelims) integral. In fact, $f : [a,b] \to \mathbb{R}$ is Riemann integrable if and only if f is bounded and continuous a.e.[8] Any such f is measurable and bounded, hence Lebesgue integrable, however this is overkill for obtaining measurability: If f is Riemann integrable, then f is bounded and there is an increasing sequence (ϕ_n) and decreasing sequence (ψ_n) of step functions such that $\phi_n \le f \le \psi_n$ and $\lim_{n\to\infty} \int_a^b \phi_n = \int_{[a,b]}^{\mathcal{R}} f = \lim_{n\to\infty} \int_a^b \psi_n$. Let $g = \sup_n \phi_n$ and $h = \inf_n \psi_n$. Then g and h are measurable, $g \le f \le h$ and $\int_{[a,b]}^{\mathcal{L}} (h-g) \le \lim_{n\to\infty} \int_a^b (\psi_n - \phi_n) = 0$. By Proposition 5.1(7), $g = h$ a.e. Then $f = g$ a.e., so f is (Lebesgue) measurable. By Corollary 5.2(3), f is Lebesgue integrable.

Moreover, for a Riemann integrable $f\colon [a,b] \to \mathbb{R}$,

$$\int_{[a,b]}^{\mathcal{R}} f = \sup\left\{\int_a^b \phi : \phi \text{ step}, \phi \le f\right\} \le \sup\left\{\int_a^b \phi : \phi \text{ simple}, \phi \le f\right\}$$
$$\le \int_{[a,b]}^{\mathcal{L}} f \le \inf\left\{\int_a^b \psi : \psi \text{ step}, f \le \psi\right\} = \int_{[a,b]}^{\mathcal{R}} f$$

Hence equality holds throughout, so the Lebesgue integral equals the Riemann integral.

[8] This is a very nice exercise, but off topic for us, so omitted. See Stein and Shakarchi Problem 1.6.4. The essential idea, which is useful for many prelims exercises relating to continuity is to consider the *oscilation* of a function f, $\omega_f(x) = \lim_{\delta\to 0}(\sup_{y\in(x-\delta,x+\delta)} f(y) - \inf_{y\in(x-\delta,x+\delta)} f(x))$. You can check that f is continuous at x if and only if $\omega_f(x) = 0$. So if f is continuous a.e. then for any $\epsilon > 0$ the set $A_\epsilon = \{x \in [a,b] : \omega_f(x) \ge \epsilon\}$ is null, so (using compactness, and you'll need to check it is compact) can be covered by finitely many open intervals of total length ϵ. This should help you access analysis 3, sheet 2, Q4 to get Riemann integrability. Conversely if f is Riemann integrable, given $n \in \mathbb{N}$ and $\epsilon > 0$, take a partition P such that $U(f;p) - L(f;P) < \epsilon/n$ and consider the total length of the intervals in P whose interior intersects $A_{1/n}$.

Given a function $f : I \to \mathbb{R}$, where I is an interval in $\mathbb{R}$, how does one test whether f is integrable over I? We can do the following:

- Note that f is measurable (for example, using Examples 3.6).
- Replace f by $|f|$: we can assume that f is non-negative. (Proposition 5.1(1),(2))
- If I is bounded and f is bounded, then f is integrable over I. (Corollary 5.2(3))
- If I or f is unbounded, we can probably consider an increasing sequence of bounded subintervals I_n, with union I, such that f is bounded on each I_n.
- We may be able to evaluate $\int_{I_n} f$ by means of the FTC, Integration by Parts, or Substitution from Prelims theory. Then we can use Baby MCT (Corollary 4.3).
- If we cannot easily evaluate the integral of f, use the Comparison Test—we look for a simpler measurable function g such that g is known to be integrable and $0 \le f \le g$ (if we think f is going to be integrable), or g is known not to be integrable and $0 \le g \le f$ (if we think f will not be integrable).

Examples 5.3. 1. Consider x^α over $(0,1)$, where $\alpha \in \mathbb{R}$. Note first that x^α is continuous, hence measurable, and non-negative. If $\alpha \ge 0$, then x^α is bounded (by 1) on $(0,1)$, hence integrable. If $\alpha < 0$, x^α has a singularity at $x = 0$, so we use Baby MCT with $I_n = [1/n, 1]$. By FTC,

$$\int_{1/n}^1 x^\alpha\,dx = \begin{cases} \frac{1-n^{-(\alpha+1)}}{\alpha+1} & (\alpha \ne -1) \\ \log n & (\alpha = -1) \end{cases} \to \begin{cases} \infty & (\alpha \le -1) \\ \frac{1}{\alpha+1} & (\alpha > -1). \end{cases}$$

By Baby MCT, x^α is integrable over $(0,1)$ if and only if $\alpha > -1$, and then $\int_0^1 x^\alpha\,dx = (\alpha+1)^{-1}$.

2. Consider x^α over $[1,\infty)$. This is similar, but with $I_n = [1,n]$. Now

$$\int_1^n x^\alpha\,dx = \begin{cases} \frac{n^{\alpha+1}-1}{\alpha+1} & (\alpha \ne -1) \\ \log n & (\alpha = -1) \end{cases} \to \begin{cases} \infty & (\alpha \ge -1) \\ -\frac{1}{\alpha+1} & (\alpha < -1). \end{cases}$$

By Baby MCT, x^α is integrable over $(1,\infty)$ if and only if $\alpha < -1$, and then $\int_1^\infty x^\alpha\,dx = -(\alpha+1)^{-1}$.

3. Consider $f(x) = x^\alpha/(1+x^\beta)$ over $(0,\infty)$, where $\alpha \in \mathbb{R}$ and $\beta \ge 0$. For $0 < x \le 1$, $x^\alpha/2 \le f(x) \le x^\alpha$. By comparison, f is integrable over $(0,1)$ if and only if x^α is, i.e., $\alpha > -1$. For $x > 1$, $x^{\alpha-\beta}/2 < f(x) < x^{\alpha-\beta}$, so, by comparison, f is integrable over $(1,\infty)$ if and only if $x^{\alpha-\beta}$ is, i.e., $\alpha - \beta < -1$. Hence f is integrable over $(0,\infty)$ if and only if $-1 < \alpha < \beta - 1$. [The case when $\beta < 0$ can be reduced to the previous case because $f(x) = x^{\alpha-\beta}/(1+x^{-\beta})$.]

4. Consider $f(x) = (\sin x)/x$ over $(0,2\pi)$. This function is continuous on $(0,2\pi]$, hence measurable. If we define $f(0) = 1$, it becomes continuous, hence bounded on $[0,2\pi]$—in fact it is bounded above by 1 and below by $-1/\pi$. So it is integrable over $(0,2\pi)$.

5. Consider $f(x) = (\sin x)/x$ over $(0,\infty)$. Now

$$\int_{r\pi}^{(r+1)\pi} \left|\frac{\sin x}{x}\right| dx \ge \int_{r\pi}^{(r+1)\pi} \frac{|\sin x|}{(r+1)\pi}\,dx = \frac{2}{(r+1)\pi}.$$

Hence,

$$\lim_{n\to\infty} \int_0^{n\pi} |f(x)|\,dx \ge \lim_{n\to\infty} \sum_{r=0}^{n-1} \frac{2}{(r+1)\pi} = \infty.$$

So $|f|$ is not integrable, and hence f is not integrable, over $(0,\infty)$.

Let us discuss the first-year theorems a little more carefully.

Theorem 5.4. (Fundamental Theorem of Calculus) *Let g be a function with a continuous derivative on a closed bounded interval $[a,b]$. Then g' is integrable over $[a,b]$, and*

$$\int_a^b g'(x)\,dx = g(b) - g(a).$$

The FTC should be treated with care, if the range of integration is unbounded (as already discussed), or if the derivative does not exist at some points as the following examples show.

Examples 5.5. 1. Let $f(x) = x\sin\left(\frac{1}{x}\right)$ $(x \in (0,1]); f(0) = 0$. Then f is continuous on $[0,1]$ and differentiable on $(0,1]$ but $f'(x) = \sin\left(\frac{1}{x}\right) - \frac{1}{x}\cos\left(\frac{1}{x}\right) \notin \mathcal{L}^1(0,1)$.

2. We define a function $\Phi : [0,1] \to [0,1]$ as follows. On the Cantor set C,

$$\Phi\left(\sum_{n=1}^{\infty} a_n 3^{-n}\right) = \sum_{n=1}^{\infty} \frac{a_n}{2} 2^{-n} \quad (a_n = 0 \text{ or } 2).$$

Then put $\Phi = \frac{1}{2}$ on $[\frac{1}{3},\frac{2}{3}]$, $\frac{1}{4}$ on $[\frac{1}{9},\frac{2}{9}]$, etc. Then Φ is continuous, monotonic, differentiable at each point of $[0,1]\setminus C$ with $\Phi'(x) = 0$. So

$$\int_0^1 \Phi'(x)\,dx = 0 \neq \Phi(1) - \Phi(0).$$

This function Φ is called the *Cantor-Lebesgue* function, or the *devil's staircase.*

Theorem 5.6. (Integration by Parts) *Let f and g be continuously differentiable functions on a closed bounded interval $[a,b]$. Then*

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx.$$

Integration by parts must be treated with great care if the interval of integration is an unbounded interval or the integrand has a singularity and you do not know whether the integrals exist. In those circumstances you cannot infer the existence of one integral from the existence of the other.

Example 5.7. Consider $\int_0^a \frac{\sin x}{x}\,dx$. Integration by parts gives

$$\int_1^a \frac{\sin x}{x}\,dx = \cos 1 - \frac{\cos a}{a} - \int_1^a \frac{\cos x}{x^2}\,dx.$$

But $\left|\frac{\cos x}{x^2}\right| \leq \frac{1}{x^2}$, so $\frac{\cos x}{x^2}$ is integrable over $[1,\infty)$, by Example 5.3(2) and the Comparison Test. It follows from Proposition 5.1(8) that

$$\lim_{a\to\infty}\int_0^a \frac{\sin x}{x}\,dx = \int_0^1 \frac{\sin x}{x}\,dx + \cos 1 - \int_1^\infty \frac{\cos x}{x^2}\,dx.$$

Nevertheless, $\sin x/x$ is not integrable over $(0,\infty)$, by Example 5.3(5).

In the case of substitution, one can infer the existence of one integral from the other. [Note: Priestley's comment near the bottom of p.133 is misleading.]

Theorem 5.8. (Substitution) *Let $g : I \to \mathbb{R}$ be a monotonic function with a continuous derivative on an interval I, and let J be the interval $g(I)$. A (measurable) function $f : J \to \mathbb{R}$ is integrable over J if and only if $(f \circ g).g'$ is integrable over I. Then*

$$\int_J f(x)\,dx = \int_I f(g(y))|g'(y)|\,dy.$$

This theorem is not contained in the one in the first-year course, because f is not required to be continuous or Riemann integrable. FTC gives the result when $f = \chi_{J'}$ for a bounded interval $J' \subseteq J$. One has to extend this to $f = \chi_E$ when $E \in \mathcal{M}_{\text{Leb}}, E \subseteq J$, i.e., one needs $m(E) = \int_{g^{-1}(E)} g'$. After that, the rest follows fairly easily. See Theorem 7.4 in Qian's notes.

Example 5.9. Let $I = (0, 1)$, $g(y) = 1/y$, so $J = (1, \infty)$. Let $f(x) = x^\alpha$. Then $x^\alpha \in \mathcal{L}^1(1, \infty)$ if and only if $y^{-\alpha-2} \in \mathcal{L}^1(0, 1)$. This provides a passage between Example 5.3, (1) and (2).

Other measures. We make some comments about integration with respect to measures other than Lebesgue.

A function $f : \mathbb{N} \to \mathbb{R}$ is integrable with respect to counting measure μ if and only if $\sum f(n)$ is absolutely convergent, and then $\int f\,d\mu = \sum_{n=1}^\infty f(n)$. Thus the general theorems that follow will provide theorems about summing absolutely convergent series.

Next, consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. A measurable function is now just a random variable X on this space, and the integral of X with respect to $\mathbb{P}$ is just the expectation $\mathbb{E}(X)$; X is integrable if and only if $|X|$ has finite expectation. The theory that follows applies to all random variables simultaneously—discrete, continuous, hybrid, singular.

6. The Convergence Theorems

The feature of Lebesgue integration theory which distinguishes it from other theories, and makes it much more manageable, is the group of theorems known as convergence theorems. These are the theorems, mentioned in the introduction, which enable one to pass limits or infinite sums through integrals, under certain conditions.

We have already seen the MCT, but we give a different form below to allow for increasing sequences of functions which are not necessarily non-negative. Notice that in this case, we must work with integrable functions so that we can add $\int f_1$ to both sides at the end of the argument.[9].

Theorem 6.1. *Let (f_n) be a sequence of integrable functions such that:*

(1) *for each n, $f_n \le f_{n+1}$ a.e.,*
(2) $\sup_n \int f_n < \infty$.

[9]Whereas the MCT which has non-negative f_n, doesn't require the f_n to have finite integrals, but then it does not conclude that f is integrable.

Then (f_n) converges a.e. to an integrable function f, and $\int f = \lim_{n\to\infty} \int f_n$.

Proof. By Proposition 5.1(6), $f_n(x) \in \mathbb{R}$ a.e. From this and assumption (1) we may redefine f_n on the union of countably many null set without changing any integrals, so we may assume that $f_n(x) \le f_{n+1}(x)$ and $f_n(x) \in \mathbb{R}$ for all x and all n. Apply Theorem 4.2 applied to $f_n - f_1$. One obtains that $\int (f - f_1) = \lim_{n\to\infty} \int f_n - \int f_1$. Thus $f - f_1$ is integrable, so f is integrable which implies that f is finite a.e. Adding $\int f_1$ to both sides we obtain that $\int f = \lim_{n\to\infty} \int f_n$. □

Theorem 6.2. [Fatou's Lemma] *Let (f_n) be a sequence of non-negative measurable functions. Then*

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n.$$

Proof. Let $g_r := \inf_{n\ge r} f_n$. Then (g_r) increases to $\liminf_{n\to\infty} f_n$ and $g_r \le f_r$ and $\int g_r \le \int f_r$. By MCT, $\int \liminf_{n\to\infty} f_n = \lim_{r\to\infty} \int g_r = \liminf_{r\to\infty} \int g_r \le \liminf_{r\to\infty} \int f_r$. □

Note that in Example 0.1 with $f_n(x) = n^2 x^n (1-x)$ on $(0,1)$, $f_n \ge 0$, $\lim_{n\to\infty} f_n = 0$, $\lim_{n\to\infty} \int f_n = 1$. So one can have $\int \limsup_{n\to\infty} f_n < \liminf_{n\to\infty} \int f_n$. However if $f_n \le g$ for all n where g is integrable, then $\int \limsup_{n\to\infty} f_n \ge \limsup_{n\to\infty} \int f_n$ (apply Fatou to $g - f_n$).

One can also have $\int \limsup_{n\to\infty} f_n > \limsup_{n\to\infty} \int f_n$—for example, $f_n(x) = \sin^2(x+n)$ on $(0,\pi)$.

Theorem 6.3. [Dominated Convergence Theorem] *Let (f_n) be a sequence of measurable functions such that:*

(1) $(f_n(x))$ converges a.e. to a limit $f(x)$,
(2) there is an integrable function g such that, for each n, $|f_n(x)| \le g(x)$ a.e.

Then f is integrable, and $\int f = \lim_{n\to\infty} \int f_n$.

Proof. Since f is measurable (Proposition 3.8) and $|f(x)| \le g(x)$ a.e., f is integrable by comparison. Apply Fatou's Lemma to $g + f_n$ and $g - f_n$, to obtain $\int (g + f) \le \int g + \liminf_{n\to\infty} \int f_n$ and $\int (g - f) \le \int g - \limsup_{n\to\infty} \int f_n$. □

Example 6.4. Consider $\displaystyle\int_0^1 \frac{n^{3/2} x e^x}{1 + n^2 x^2}\, dx$. It is difficult (impossible?) to evaluate the integrals themselves, but we can find the limit of the integrals, with the help of the DCT (Theorem 6.3). Let

$$f_n(x) = \frac{n^{3/2} x e^x}{1 + n^2 x^2} = \frac{(nx)^{3/2}}{1 + n^2 x^2} \frac{e^x}{x^{1/2}}.$$

The function $\displaystyle\frac{y^{3/2}}{1+y^2}$ tends to 0 as $y \to \infty$, so it is bounded for $y > 0$. It follows that $f_n(x) \to 0$ as $n \to \infty$, and there is a constant c such that

$$0 \le f_n(x) \le \frac{ce^x}{x^{1/2}} \le \frac{ce}{x^{1/2}} \quad (0 < x < 1).$$

Now let $g(x) = \dfrac{ce}{x^{1/2}}$. Then g is integrable over $(0,1)$ (Example 5.3(1)), so we have verified the conditions of the DCT (with $f = 0$). We can therefore conclude that

$$\lim_{n\to\infty} \int_0^1 \frac{n^{3/2} x e^x}{1 + n^2 x^2}\, dx = 0.$$

Corollary 6.5. [Bounded Convergence Theorem] *Let I be a bounded interval, (f_n) be a sequence in $\mathcal{L}^1(I)$ converging a.e. to f, and suppose that there is a constant c such that $|f_n(x)| \le c$ a.e., for all n. Then $f \in \mathcal{L}^1(I)$, and $\int_I f = \lim_{n\to\infty} \int_I f_n$.*

The next example involves, for the first time in this course, integration of a complex-valued function. A function $f : \mathbb{R} \to \mathbb{C}$ is *integrable* if $\operatorname{Re} f$ and $\operatorname{Im} f$ are both integrable. Results which hold for real-valued integrable functions and which make sense for complex-valued functions are almost invariably true in the complex case, and can easily be deduced by applying the result to the real and imaginary parts separately. This is the case, for example, with the Comparison Test, FTC, Integration by Parts and the DCT. Note, however, that in Theorem 5.8 (Substitution), the function f may be complex-valued, but the substitution $g(t)$ is assumed to be real-valued.

Example 6.6. Let γ_r be the semi-circular contour $\{re^{i\theta} : 0 \le \theta \le \pi\}$, and consider

$$\int_{\gamma_r} \frac{e^{iz}}{z}\, dz = i \int_0^\pi e^{ir\cos\theta} e^{-r\sin\theta}\, d\theta.$$

Since

$$\begin{aligned} \left| e^{ir\cos\theta} e^{-r\sin\theta} \right| &\le 1 \quad \text{for all } r > 0,\ 0 \le \theta \le \pi \\ e^{ir\cos\theta} e^{-r\sin\theta} &\to \begin{cases} 0 & \text{as } r \to \infty, \text{ if } 0 < \theta < \pi, \\ 1 & \text{as } r \to 0 \end{cases} \end{aligned}$$

the Bounded Convergence Theorem gives

$$\int_{\gamma_{R_n}} \frac{e^{iz}}{z}\, dz \to 0 \quad (R_n \to \infty), \qquad \int_{\gamma_{\varepsilon_n}} \frac{e^{iz}}{z}\, dz \to \pi i \quad (\varepsilon_n \to 0).$$

By Cauchy's Theorem,

$$0 = \int_{\gamma_{R_n}} \frac{e^{iz}}{z}\, dz - \int_{\gamma_{\varepsilon_n}} \frac{e^{iz}}{z}\, dz + \int_{\varepsilon_n}^{R_n} \frac{e^{ix} - e^{-ix}}{x}\, dx.$$

Letting $n \to \infty$, we obtain

$$\lim_{n\to\infty} \int_{\varepsilon_n}^{R_n} \frac{\sin x}{x}\, dx = \frac{\pi}{2}.$$

Hence $\lim_{a\to\infty} \int_0^a \frac{\sin x}{x}\, dx = \pi/2$ (see Example 5.7, and Part A Complex Analysis, Example 11.9 in MT2020 notes).

Next we will apply the results above to term-by-term integration of series. We start by recalling the MCT for Series (Corollary 4.5 above).

Theorem 6.7. [Monotone Convergence Theorem for Series] *Let (g_n) be a sequence of integrable functions such that:*

(1) for each n, $g_n \geq 0$ a.e.,
(2) $\sum_n \int g_n < \infty$.

Then $\sum_{n=1}^\infty g_n$ converges a.e. to an integrable function, and $\int \sum_{n=1}^\infty g_n = \sum_{n=1}^\infty \int g_n$.

Theorem 6.8. [Lebesgue's Series Theorem; Beppo Levi Theorem, ...] *Let (g_n) be a sequence of integrable functions such that $\sum_n \int |g_n| < \infty$. Then $\sum_{n=1}^\infty g_n$ converges a.e. to an integrable function, and $\int \sum_{n=1}^\infty g_n = \sum_{n=1}^\infty \int g_n$.*

Proof. Apply MCT for Series to g_n^+ and g_n^-. Alternatively, apply MCT for Series to $|g_n|$ and use the fact that absolute convergence implies convergence. □

Theorem 6.9. *Let (g_n) be a sequence of integrable functions such that $\sum_n |g_n|$ is integrable. Then $\sum_{n=1}^\infty g_n$ converges a.e. to an integrable function, and $\int \sum_{n=1}^\infty g_n = \sum_{n=1}^\infty \int g_n$.*

Proof. Clearly $\sum_{n=1}^k \int |g_n| \leq \int \sum_{n=1}^\infty |g_n|$ for all k, so $\sum_{n=1}^\infty \int |g_n| \leq \int \sum_{n=1}^\infty |g_n|$. Apply Theorem 6.8. □

Example 6.10. Let $\alpha > 0$, and consider $\int_0^1 x^{\alpha-1} e^{-x}\, dx$. Let $g_n(x) = (-1)^n x^{\alpha+n-1}/n!$, so that $\sum_{n=0}^\infty g_n(x) = x^{\alpha-1}e^{-x}$. Now

$$\int_0^1 |g_n(x)|\, dx = \frac{1}{(\alpha+n)n!},$$

so $\sum_n \int_0^1 |g_n(x)|\, dx < \infty$. Thus Lebesgue's Series Theorem tells us that our integral exists (we could have established this directly, by comparing the integrand with $x^{\alpha-1}$), and that

$$\int_0^1 x^{\alpha-1}e^{-x}\, dx = \sum_{n=0}^\infty \int_0^1 (-1)^n x^{\alpha+n-1}/n!\, dx = \sum_{n=0}^\infty \frac{(-1)^n}{(\alpha+n)n!}.$$

Example 6.11. Let $s \in \mathbb{R}$, and consider $\int_{-\infty}^\infty e^{-isx}e^{-x^2}\, dx$. The integrand is continuous, $|e^{-isx}e^{-x^2}| = e^{-x^2} \leq ee^{-|x|} \in \mathcal{L}^1$ (exercise). If $g_n(x) = \frac{(-isx)^n}{n!}e^{-x^2}$, then $\sum_{n=0}^\infty g_n(x) = e^{-isx}e^{-x^2}$, and

$$\sum_{n=0}^\infty |g_n(x)| = e^{|sx|-x^2} \leq e^{s^2/2}e^{-x^2/2} \in \mathcal{L}^1.$$

It follows that $\sum_n |g_n| \in \mathcal{L}^1$, so Theorem 6.9 shows that term-by-term integration is permissible, and

$$\int_{-\infty}^\infty e^{-isx}e^{-x^2}\, dx = \sum_{n=0}^\infty \int_{-\infty}^\infty \frac{(-isx)^n}{n!} e^{-x^2}\, dx.$$

Now

$$\int_{-\infty}^\infty x^n e^{-x^2}\, dx = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \dfrac{(2m)!\sqrt{\pi}}{4^m m!} & \text{if } n = 2m, \end{cases}$$

(for $m = 0$ this is a standard trick, and one can use integration by parts and induction on m). Thus

$$\int_{-\infty}^{\infty} e^{-isx} e^{-x^2}\,dx = \sum_{m=0}^{\infty} \frac{(-is)^{2m}\sqrt{\pi}}{4^m m!} = \sqrt{\pi}e^{-s^2/4}.$$

The integral which we have just evaluated is very important—for example, apart from a few constants, it is the characteristic function of the normal distribution (as in Part A Probability); in analysts' language, it is the Fourier transform of the function e^{-x^2} (as in DEs). There are other methods of evaluating the integral; one is given in Priestley (*Complex Analysis*, 22.12) and Part A Integral Transforms (Example 77 in HT2020 notes), and another will be given in Example 7.6.

All theorems in this Section hold in general measure spaces. Corollary 6.5 holds in finite measure spaces.

7. Integrals depending on a parameter

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function of two variables. In a while, we shall discuss the (double) integral, and the repeated integrals, of f. First, we merely consider the partial integral of f, obtained by integration with respect to one of the variables. Thus we suppose that for each fixed y, the function $x \mapsto f(x, y)$ is integrable. We can then define a function F by:

$$F(y) = \int f(x, y)\,dx.$$

A natural, and important, question is whether F is continuous, or differentiable, assuming that f has corresponding properties. In general, the answer is negative (see Example 7.1), but if we impose some mild conditions of the type that appear in the DCT, then the answer is positive.

Example 7.1. Let $f(x, y) = ye^{-x^2y^2}$. Since $f(x, 0) = 0$ for all x, $F(0) = 0$. For fixed $y \neq 0$, we can make the substitution $t = yx$ and deduce that $F(y) = \int_{-\infty}^{\infty} e^{-t^2}\,dt (= \sqrt{\pi})$ $(y \neq 0)$. Thus F is discontinuous, even though f is differentiable.

Theorem 7.2. [Continuous-parameter DCT] *Let I and J be intervals in $\mathbb{R}$, and $f : I \times J \to \mathbb{R}$ be a function such that:*

(1) for each y in J, $x \mapsto f(x, y)$ is integrable over I,
(2) for each y in J, $\lim_{y' \to y} f(x, y') = f(x, y)$ a.e.(x),
(3) there exists an integrable function g on I such that for each y in J, $|f(x, y)| \leq g(x)$ a.e.(x).

Define $F(y) = \int_I f(x, y)\,dx \quad (y \in J)$. Then F is continuous on J.

Remark. In condition (3) of Theorem 7.2, *the function g does not depend on y.*

Proof. Let (y_n) be any sequence in J converging to $y \in J$. Let $f_n(x) = f(x, y_n)$. Then $|f_n(x)| \leq g(x)$ a.e., for all n, and $\lim_{n\to\infty} f_n(x) = f(x, y)$ a.e., so the conditions of the

DCT are satisfied. The DCT implies that:

$$F(y_n) = \int_I f(x, y_n)\, dx \to \int_I f(x, y)\, dx = F(y).$$

Thus F is continuous. □

Example 7.3. The Gamma function Γ is defined by:

$$\Gamma(y) = \int_0^\infty e^{-x} x^{y-1}\, dx \quad (y > 0).$$

We wish to show that Γ is continuous, firstly for $y \in [1, 2]$. In order to apply Theorem 7.2, we take $I = (0, \infty)$, $J = [1, 2]$, and $f(x, y) = e^{-x} x^{y-1}$. Condition (1) of Theorem 7.2 is an exercise, and (2) is more or less trivial. For condition (3), we need to ensure that

$$g(x) \geq \sup_{1 \leq y \leq 2} f(x, y) = \begin{cases} e^{-x} & (0 < x \leq 1) \\ xe^{-x} & (x > 1). \end{cases} \tag{7.1}$$

We choose to take g *equal* to the right-hand side of (7.1). Then g is integrable over $(0, \infty)$ (exercise), so condition (3) of Theorem 7.2 is satisfied. Thus, Theorem 7.2 shows that Γ is continuous on $[1, 2]$.

In fact, Γ is continuous on $(0, \infty)$. However, it is impossible to establish this by applying Theorem 7.2 with $J = (0, \infty)$, for in condition (3), it would be necessary that

$$g(x) \geq \sup_{y > 0} f(x, y) = \begin{cases} x^{-1} e^{-x} & (0 < x \leq 1) \\ \infty & (x > 1). \end{cases}$$

Such a function g cannot possibly be integrable over $(0, \infty)$, so it is impossible to satisfy condition (3) of Theorem 7.2. Instead, we proceed as follows. For each $b > 0$, let $J_b = (a, c)$, where a and c are chosen so that $0 < a < b < c$, for example, $a = b/2$ and $c = 2b$. Then let

$$g_b(x) = \sup_{a < y < c} f(x, y) = \begin{cases} x^{a-1} e^{-x} & (0 < x \leq 1) \\ x^{c-1} e^{-x} & (x > 1). \end{cases}$$

Then g_b is integrable over $(0, \infty)$. Thus, Theorem 7.2 shows that Γ is continuous on (a, c), and in particular at b. But b is arbitrary, so Γ is continuous on $(0, \infty)$.

The point is that continuity is a local property: F is continuous if F is continuous at y for all y in the domain. We abstract the method to obtain the following version of Theorem 7.2, where the dominating function is defined locally depending on the parameter. Notice though that we still need a single g_b to be valid over the entire open interval J_b.

Corollary 7.4. *Let I and J be intervals in $\mathbb{R}$, and $f : I \times J \to \mathbb{R}$ be a function such that (1) and (2) of Theorem 7.2 hold, and*

(3′) *for each $b \in J$, there exist an open subinterval J_b of J containing b and an integrable function g_b on I such that, for each $y \in J_b$, $|f(x, y)| \leq g_b(x)$ a.e.(x).*

Then F is continuous on J, where F is as in Theorem 7.2.

Remark. The method of Theorem 7.2 can also be used to cover cases where $y \to y_0$ for a single point y_0 or $y \to \infty$. For example, suppose that there exists a in $\mathbb{R}$ and a function $h : I \to \mathbb{R}$ such that

(1) for each $y > a$, $x \mapsto f(x,y)$ is integrable over I,
(2) $\lim_{y\to\infty} f(x,y) = h(x)$ a.e.(x),
(3) there exists an integrable function g on I such that for each $y > a$, $|f(x,y)| \leq g(x)$ a.e.(x).

Then $F(y) \to \int_I h(x)\,dx$ as $y \to \infty$.

Now we turn to the question of differentiability of F. The sort of result which we hope to have is that if $\frac{\partial f}{\partial y}$ exists, and some supplementary conditions are satisfied, then F is differentiable and

$$F'(y) = \int \frac{\partial f}{\partial y}(x,y)\,dx$$

(differentiation through, or under, the integral sign). The standard supplementary condition is that $\frac{\partial f}{\partial y}$ should be dominated by an integrable function, *independent of* y.

Theorem 7.5. *Let I and J be intervals in $\mathbb{R}$, and $f : I \times J \to \mathbb{R}$ be a function such that:*

(1) for each y in J, $x \mapsto f(x,y)$ is integrable over I,
(2) for almost all x in I, $\frac{\partial f}{\partial y}(x,y)$ exists for all $y \in J$
(3) there is an integrable function $g : I \to \mathbb{R}$ such that for almost all $x \in I$, $\left|\frac{\partial f}{\partial y}(x,y)\right| \leq g(x)$ holds for all $y \in J$.

Define $F(y) = \int_I f(x,y)\,dx \quad (y \in J)$. Then F is differentiable on J and

$$F'(y) = \int_I \frac{\partial f}{\partial y}(x,y)\,dx.$$

Note that in condition (3) (and (2)) above we require a single null set N such that $\left|\frac{\partial f}{\partial y}(x,y)\right| \leq g(x)$ holds for all $y \in J$ and $x \in I \setminus N$. This is not a-priori the same as requiring that for all $y \in J$, $\left|\frac{\partial f}{\partial y}(x,y)\right| \leq g(x)$ holds for almost all $x \in I$.[10] In practise when you want to apply Theorem 7.5 it's quite likely that (3) will hold for all x and y (or perhaps all but finitely many values of x).

Proof. Fix y in J, and let (y_n) be any sequence in J converging to y (with $y_n \neq y$). Let

$$g_n(x) = \frac{f(x,y_n) - f(x,y)}{y_n - y}.$$

Then g_n is integrable over I, $g_n(x) \to \frac{\partial f}{\partial y}(x,y)$ for almost all x as $n \to \infty$. Moreover, the Mean Value Theorem says that there exists a point $\xi_{x,n}$ (depending on x and n)

[10] as that would allow the null set N_y of those x for which this fails to depend on y.

between y_n and y such that $g_n(x) = \frac{\partial f}{\partial y}(x, \xi_{x,n})$. It follows from (3) that $|g_n(x)| \le g(x)$ a.e.(x).[11] This shows that the Dominated Convergence Theorem is applicable, so

$$\frac{F(y_n) - F(y)}{y_n - y} = \int_I g_n(x)\, dx \to \int_I \frac{\partial f}{\partial y}(x, y)\, dx \quad \text{as } n \to \infty.$$

Since (y_n) is an arbitrary sequence tending to y, and the right-hand side is independent of the choice of sequence, it follows that

$$\frac{F(y') - F(y)}{y' - y} \to \int_I \frac{\partial f}{\partial y}(x, y)\, dx \quad \text{as } y' \to y,$$

which completes the proof. □

Example 7.6. Let $f(x, s) = e^{-isx}e^{-x^2}$, and $F(s) = \int_{-\infty}^{\infty} f(x, s)\, dx$ (compare Example 6.11). This integral exists for all s. Moreover,

$$\frac{\partial f}{\partial s}(x, s) = -ixe^{-ixs}e^{-x^2},$$

so

$$\left|\frac{\partial f}{\partial s}(x, s)\right| = |x|e^{-x^2}.$$

Since

$$\int_{-n}^{n} |x|e^{-x^2}\, dx = 2\int_0^n xe^{-x^2}\, dx = 1 - e^{-n^2} \to 1$$

as $n \to \infty$, $|x|e^{-x^2} \in \mathcal{L}^1(\mathbb{R})$ (Baby MCT). Thus Theorem 7.5 is applicable, with $I = J = \mathbb{R}$ and $g(x) = |x|e^{-x^2}$. It follows that F is differentiable on $\mathbb{R}$, and

$$F'(s) = -i\int_{-\infty}^{\infty} xe^{-isx}e^{-x^2}\, dx.$$

By integration by parts,

$$F'(s) = -\frac{s}{2}F(s).$$

Hence $F(s) = Ae^{-s^2/4}$ for some constant A. But $F(0) = \int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$, so $A = \sqrt{\pi}$.

Corollary 7.7. *Let I and J be intervals in $\mathbb{R}$, and $f : I \times J \to \mathbb{R}$ be a function such that (1) and (2) of Theorem 7.5 hold, and*

(3′) *for each b in J, there is an open subinterval J_b of J containing b and an integrable function $g_b : I \to \mathbb{R}$ such that, for almost all x, $\left|\frac{\partial f}{\partial y}(x, y)\right| \le g_b(x)$ for all $y \in J_b$.*

Then the conclusions of Theorem 7.5 hold.

[11] This is where it matters that we have a single null set N for which $\left|\frac{\partial f}{\partial y}(x, y)\right| \le g(x)$ holds for all $x \in I \setminus N$ and $y \in J$. If we had that for each $y \in J$, there was a null set N_y depending on y such that the estimate holds for $y \in J$ and $x \in I \setminus N_y$, then as $\xi_{x,n}$ depends on x, we would have no way of deducing that $|g_n(x)| \le g(x)$ a.e.. I think Oliver Riordan for bringing this issue to my attention.

Example 7.8. Let $f(x,y) = e^{-xy}(1+x^3)^{-1}$ $(x \geq 0, y \geq 0)$. Since $0 \leq f(x,y) \leq (1+x^3)^{-1}$, $x \mapsto f(x,y)$ is integrable over $[0,\infty)$ for each $y \geq 0$. Moreover,

$$\frac{\partial f}{\partial y}(x,y) = -\frac{xe^{-xy}}{1+x^3},$$

so

$$\left|\frac{\partial f}{\partial y}(x,y)\right| \leq \frac{x}{1+x^3} \quad (x \geq 0, y \geq 0).$$

Since $x(1+x^3)^{-1}$ is integrable over $[0,\infty)$ (by comparison with x^{-2} for $x \geq 1$), Theorem 7.5 is applicable, and shows that F is differentiable on $[0,\infty)$ and

$$F'(y) = -\int_0^\infty \frac{xe^{-xy}}{1+x^3}\,dx.$$

We would like to repeat this argument to show that $F''(y)$ exists (at least for $y > 0$), but this is more complicated. Indeed,

$$\frac{\partial^2 f}{\partial y^2}(x,y) = \frac{x^2 e^{-xy}}{1+x^3}.$$

For $y = 0$, this function is not integrable (by comparison with x^{-1}), so we should only consider $y > 0$. However, it is not possible to apply Theorem 7.5 with f replaced by $\frac{\partial f}{\partial y}$ and with $J = (0,\infty)$, because

$$\sup_{y>0} \frac{\partial^2 f}{\partial y^2}(x,y) = \frac{x^2}{1+x^3},$$

which is not integrable over $[0,\infty)$. Instead, we must apply Corollary 7.7. Thus we take $b > 0$, let $J_b = (b/2,\infty)$, and

$$g_b(x) = \sup_{y>b/2} \frac{\partial^2 f}{\partial y^2}(x,y) = \frac{x^2 e^{-xb/2}}{1+x^3} \leq x^2 e^{-xb/2}.$$

This function is integrable on $[0,\infty)$, and we conclude from Corollary 7.7, with f replaced by $\frac{\partial f}{\partial y}$ and $J = (0,\infty)$ that $F''(y)$ exists for $y > 0$ and

$$F''(y) = \int_0^\infty \frac{x^2 e^{-xy}}{1+x^3}\,dx.$$

Repeating this argument, it is possible to show that F is infinitely differentiable on $(0,\infty)$ and to obtain integrals for all the derivatives.

Remark. There are versions of Theorem 7.5 and Corollary 7.7 where the real variable $y \in J$ is replaced by a complex variable $z \in \Omega$, a domain in $\mathbb{C}$, the function f is complex-valued, $z \mapsto f(x,z)$ is holomorphic for each x, and the conclusion is that F is holomorphic. The proofs are almost the same, except that the use of the Mean Value Theorem should be replaced by the formula $g_n(x) = (z_n - z_0)^{-1} \int_{[z_0,z_n]} \frac{\partial f}{\partial w}(x,w)\,dw$ which leads to the estimate $|g_n(x)| \leq g(x)$.

8. Double Integrals

In Section 7, we considered some properties concerning functions of two variables, but we confined integration to one of the variables. Now it is time to consider integration with respect to both variables. An example on Problem Sheet 1 shows that this is not just a matter of integrating first with respect to one variable, and then with respect to the other (repeated integration). What one has to do is to define the class $\mathcal{L}^1(\mathbb{R}^2)$ of integrable functions on $\mathbb{R}^2$, and their (double) integrals, in a way which treats both variables simultaneously, then establish the theorem (Fubini's Theorem) which ensures that the double integrals coincide with the repeated integrals for functions in $\mathcal{L}^1(\mathbb{R}^2)$, and establish a practical method (Tonelli's Theorem) to determine whether a given function is integrable.

The first part of this is routine. The class $\mathcal{L}^1(\mathbb{R}^2)$ of integrable functions on $\mathbb{R}^2$ is defined in exactly the same way as $\mathcal{L}^1(\mathbb{R})$, except that intervals (a, b), and their lengths $b - a$, are replaced by rectangles $(a, b) \times (c, d)$ and their areas $(b - a)(d - c)$. Then one defines outer measure, null sets (line segments etc are null), measurable sets (all open sets etc are measurable), as we discussed at the end of Section 2, measurable functions, simple functions, integrable functions and (double) integrals just as in Sections 2–4. Moreover, the results of Sections 2-6, except Section 4 from Theorem 5.4 onwards, remain valid, with obvious changes of wording where necessary. More details may be found in Capinski & Kopp (Chap 6, but in greater generality) or Stein & Shakarchi (from beginning).

The (double) integral of an integrable function f over $\mathbb{R}^2$ may be denoted by any of the following:

$$\int f, \quad \int_{\mathbb{R}^2} f, \quad \int f(x, y)\, d(x, y), \quad \int_{\mathbb{R}^2} f(x, y)\, d(x, y).$$

Theorem 8.1. (Tonelli) *Let $f : \mathbb{R}^2 \to [0, \infty]$ be measurable. Then*

(1) $x \mapsto f(x, y)$ is measurable for almost all y;
(2) $y \mapsto \int_{\mathbb{R}} f(x, y)\, dx$ (defined a.e.) is measurable;
(3)

$$\int_{\mathbb{R}^2} f(x, y)\, d(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dx \right) dy.$$

Now we state two consequences of this in their traditional form.

Theorem 8.2. [Fubini's Theorem] *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be integrable. Then, for almost all y, the function $x \mapsto f(x, y)$ is integrable. Moreover, defining (for almost all y) by $F(y) = \int f(x, y)\, dx$, then F is integrable, and*

$$\int_{\mathbb{R}^2} f(x, y)\, d(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dx \right) dy.$$

Similarly,

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dx \right) dy = \int_{\mathbb{R}^2} f(x, y)\, d(x, y) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y)\, dy \right) dx,$$

where the first repeated integral exists in the sense described above.

Proof. Apply Theorem 8.1 to f^+ and f^-, using Proposition 5.1(6) to get that $\int_{\mathbb{R}} f^{\pm}(x,y)\,dx < \infty$ a.e.(y). $\square$

Theorem 8.3. [Tonelli's Theorem] *Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a measurable function, and suppose that either of the following repeated integrals is finite:*

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|\,dx \right) dy, \qquad \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |f(x,y)|\,dy \right) dx.$$

Then f is integrable. Hence, Fubini's Theorem is applicable to both f and $|f|$.

Proof. Apply Theorem 8.1 to get that $\int_{\mathbb{R}^2} |f| < \infty$. Then $f \in \mathcal{L}^1(\mathbb{R}^2)$, by Proposition 5.1(2). $\square$

Remark. Note that, when applying Tonelli's Theorem, one must verify that a repeated integral of $|f|$ is finite. It is not sufficient that the repeated integrals of f exist (see Example 8.4), nor is it sufficient that the repeated integrals of f both exist and are equal (see Example 8.7).

If E is a measurable subset of $\mathbb{R}^2$ and $f : E \to \mathbb{R}$ is any function, then f is said to be integrable over E if $\tilde{f}$ is integrable over $\mathbb{R}^2$, where $\tilde{f}(x,y) = f(x,y)$ if $(x,y) \in E$, $\tilde{f}(x,y) = 0$ otherwise. Then $\int_E f$ is defined to be $\int_{\mathbb{R}^2} \tilde{f}$.

Fubini's Theorem and Tonelli's Theorem can be applied in this situation. However, when E is not a rectangle, great care must be taken to choose the correct limits of integration in the repeated integrals. If in any doubt draw a sketch of the region. See Example 8.5.

In repeated integrals, one often omits the brackets around the inner integral and writes $\iint f(x,y)\,dy\,dx$, etc., with appropriate limits of integration. This means that one is integrating first with respect to y between the limits on the right-hand integral sign, which may be functions of x. Thus

$$\int_a^b \int_{\phi(x)}^{\psi(x)} f(x,y)\,dy\,dx$$

denotes the repeated integral over the region E bounded by curves $y = \phi(x)$ and $y = \psi(x)$ and by vertical lines $x = a$, $x = b$.

Example 8.4. Let $f(x,y) = \dfrac{x-y}{(x+y)^3}$ $\quad (0 < x < 1, 0 < y < 1)$. It was an exercise in Problem Sheet 1 that the repeated integrals of f exist, but are not equal. It follows from the final part of Fubini's Theorem that f is not integrable over the square $(0,1) \times (0,1)$.

Example 8.5. Consider $\displaystyle\int_0^1 \left(\int_0^x \left(\frac{1-y}{x-y} \right)^{1/2} dy \right) dx$. As it stands, the inner integral is difficult. However, it turns out that when the order of integration is reversed, the other repeated integral is easily evaluated. To justify the equality of the repeated integrals, we apply Tonelli's Theorem; this is contained in the following discussion.

First, note that the integrand is continuous except on the line $y = x$ which is null; it is non-negative throughout the range of integration, so that in applying Tonelli's Theorem, it is unnecessary to replace f by $|f|$. The next problem is to work out the limits of integration when we reverse the order. For this, we have to identify the region in $\mathbb{R}^2$ over which the double integral is taken. For each x, between 0 and 1, we are integrating along the (vertical) line-segment from $y = 0$ to $y = x$. As x runs from 0 to 1, this sweeps out the triangle shown. The integrand is continuous on the interior of the triangle (and we take it to be 0 outside the triangle), so it is measurable. If we fix a value of y, the values of x which give us points within the triangle are those between $x = y$ and $x = 1$. This applies for y between 0 and 1; otherwise there are no points within the triangle. Thus the limits of the reversed repeated integral are $x = y$ and $x = 1$ in the inner integral, and $y = 0$ and $y = 1$ in the outer. This is confirmed by the following equalities of sets:

$$\begin{aligned}\{(x,y) \in \mathbb{R}^2 : 0 < y < x, 0 < x < 1\} &= \{(x,y) \in \mathbb{R}^2 : 0 < y < x < 1\} \\ &= \{(x,y) \in \mathbb{R}^2 : y < x < 1, 0 < y < 1\},\end{aligned}$$

but the picture was more informative!

Now the reversed repeated integral is:

$$\begin{aligned}\int_0^1 \left(\int_y^1 \left(\frac{1-y}{x-y} \right)^{1/2} dx \right) dy &= \int_0^1 \left[2(1-y)^{1/2}(x-y)^{1/2} \right]_{x=y}^{x=1} dy \\ &= \int_0^1 2(1-y)\, dy = 1.\end{aligned}$$

Since the integrand is non-negative, and since this repeated integral is finite, it follows from Tonelli's Theorem that f is integrable over the triangle, and from Fubini's Theorem that

$$\int_0^1 \left(\int_0^x \left(\frac{1-y}{x-y} \right)^{1/2} dy \right) dx = 1.$$

The next example shows how it is both possible and useful to make changes of variable within the inner integral of a repeated integral. The same technique will be used in several subsequent examples.

Example 8.6. Let $f(x,y) = ye^{-y^2(1+x^2)}$. Since f is continuous, it is certainly measurable. We shall consider the integral of f over the positive quadrant $(0,\infty)\times(0,\infty)$. First we consider $\int_0^\infty f(x,y)\,dy$ for a fixed x. Making the change of variable $t = y(1+x^2)^{1/2}$ (x is a constant at this point),

$$\int_0^\infty f(x,y)\,dy = \int_0^\infty \frac{te^{-t^2}}{1+x^2}\,dt = \lim_{k\to\infty} \left[-\frac{e^{-t^2}}{2(1+x^2)} \right]_{t=0}^{t=k} = \frac{1}{2(1+x^2)}.$$

This function is integrable with respect to x, and

$$\int_0^\infty \left(\int_0^\infty f(x,y)\,dy \right) dx = \frac{\pi}{4}.$$

Since $f(x,y) \geq 0$ for $y \geq 0$, it follows from Tonelli's Theorem that f is integrable over $(0,\infty) \times (0,\infty)$, and by Fubini's Theorem,

$$\int_0^\infty \left(\int_0^\infty f(x,y)\,dx \right) dy = \frac{\pi}{4}.$$

In the inner integral, where $y > 0$ is fixed, we can make the change of variable $u = xy$, and obtain

$$\begin{aligned}\frac{\pi}{4} = \int_0^\infty \left(\int_0^\infty e^{-(y^2+u^2)}\,du \right) dy &= \int_0^\infty e^{-y^2} \left(\int_0^\infty e^{-u^2}\,du \right) dy \\ &= \left(\int_0^\infty e^{-u^2}\,du \right) \left(\int_0^\infty e^{-y^2}\,dy \right) = \left(\int_0^\infty e^{-x^2}\,dx \right)^2.\end{aligned}$$

It follows that

$$\int_0^\infty e^{-x^2}\,dx = \frac{\sqrt{\pi}}{2}.$$

If f takes both positive and negative values, then to apply Tonelli's Theorem, it is necessary to consider $|f|$, or alternatively to consider separately the regions where f is positive and where it is negative.

Example 8.7. Let $f(x,y) = \dfrac{xy}{x^4+y^4}$. Since f is odd both as a function of x, and also as a function of y, $\int_{-\infty}^\infty f(x,y)\,dy = 0$ for all x, and $\int_{-\infty}^\infty f(x,y)\,dx = 0$ for all y. Hence both repeated integrals exist and equal 0. However, if we consider f over the quadrant $x > 0, y > 0$, part of the region where $f(x,y) > 0$, then, putting $y = xt$ ($x > 0$ fixed),

$$\int_0^\infty f(x,y)\,dy = \int_0^\infty \frac{x^3 t}{x^4(1+t^4)}\,dt = \frac{c}{x},$$

where c is the constant $\int_0^\infty \dfrac{t}{1+t^4}\,dt$. Since cx^{-1} is not integrable with respect to x over $(0,\infty)$, it follows that f is not integrable over the quadrant, and therefore not integrable over the plane.

In practice, it often happens that one has no means of evaluating the repeated integrals of f or $|f|$, but can nevertheless decide whether f is integrable. One technique for this is to show that f is dominated by a simpler function which one can show to be integrable (or that f dominates a function which one can show not to be integrable).

Example 8.8. Let $f(x,y) = \sin\left(\dfrac{1}{x^2+y^4}\right)\cos(x^2+y^3)$. We wish to show that f is integrable over the positive quadrant $(0,\infty)\times(0,\infty)$.
Since f is continuous in this region (although not continuous at $(0,0)$), it is measurable. Moreover, f is

bounded, and hence integrable over any bounded region, in particular over the square $(0,1)\times(0,1)$. Thus it suffices to show that f is integrable over the regions $[1,\infty)\times[0,\infty)$ and $(0,1)\times(1,\infty)$.

Using the inequalities $|\sin t| \le |t|$ and $|\cos t| \le 1$, it follows that $|f(x,y)| \le (x^2+y^4)^{-1}$, so it suffices to show that $(x^2+y^4)^{-1}$ is integrable over these two regions. Now

$$\int_1^\infty \left(\int_0^\infty \frac{dy}{x^2+y^4}\right) dx = \int_1^\infty \left(\int_0^\infty \frac{dz}{x^{3/2}(1+z^4)}\right) dx < \infty,$$

where we made the substitution $y = x^{1/2}z$ and used the integrability of $x^{-3/2}$ over $[1,\infty)$ and of $(1+z^4)^{-1}$ over $(0,\infty)$. Also,

$$\int_1^\infty \left(\int_0^1 \frac{dx}{x^2+y^4}\right) dy \le \int_1^\infty \left(\int_0^1 \frac{dx}{y^4}\right) dy = \int_1^\infty \frac{dy}{y^4} = \frac{1}{3}.$$

It follows from Tonelli's Theorem that $(x^2+y^4)^{-1}$ is integrable over these two regions, so f is integrable over the quadrant.

Another useful technique for testing functions for integrability, and for evaluating integrals, is to change variables. The reader will be familiar with this idea from courses in applied mathematics and in A3 Probability, and will know that one has to take account of the Jacobian of the transformation. The method is the extension to two variables of Theorem 5.8. We shall state the result and give examples for polar coordinates $x = r\cos\theta$, $y = r\sin\theta$, when the Jacobian is r. This corresponds to the fact that a small rectangle with sides $\delta r, \delta\theta$ (area $\delta r\delta\theta$) in the (r,θ)-space is transformed into an approximate rectangle of sides $\delta r, r\delta\theta$ (area $r\delta r\delta\theta$)) in the (x,y)-space.

Theorem 8.9. *Let E be a measurable subset of $\mathbb{R}^2$, and $f : E \to \mathbb{R}$ be a function. Let $E' = \{(r,\theta) : 0 \le r, 0 \le \theta < 2\pi, (r\cos\theta, r\sin\theta) \in E\}$ and $g(r,\theta) = rf(r\cos\theta, r\sin\theta)\,(r,\theta \in E')$. Then f is integrable over E if and only if g is integrable over E'. In that case,*

$$\int_E f(x,y)\,d(x,y) = \int_{E'} f(r\cos\theta, r\sin\theta)\,r\,d(r,\theta).$$

Example 8.10. In Example 8.6 we evaluated $\int_0^\infty e^{-x^2}\,dx$, using Fubini's Theorem. Here, we shall evaluate the same integral by the more common method of polar coordinates.

Let $E = (0,\infty)\times(0,\infty)$ and $f(x,y) = e^{-(x^2+y^2)}$. Then

$$\int_0^\infty \int_0^\infty f(x,y)\,dy\,dx = \left(\int_0^\infty e^{-x^2}\,dx\right)\left(\int_0^\infty e^{-y^2}\,dy\right) = \left(\int_0^\infty e^{-x^2}\,dx\right)^2 < \infty.$$

It follows from Tonelli's Theorem that f is integrable over E. In the notation of Theorem 8.9, $E' = \{(r,\theta) : 0 < r, 0 < \theta \le \pi/2\}$, so it follows from Theorem 8.9 and Fubini's Theorem that

$$\left(\int_0^\infty e^{-x^2}\,dx\right)^2 = \int_{E'} e^{-r^2} r\,d(r,\theta) = \int_0^{\pi/2}\int_0^\infty e^{-r^2} r\,dr\,d\theta = \frac{\pi}{4}.$$

This confirms that $\int_0^\infty e^{-x^2}\,dx = \sqrt{\pi}/2$.

Example 8.11. As in Example 8.7, let $f(x,y) = \dfrac{xy}{x^4+y^4}$. In the notation of Theorem 8.9, $g(r,\theta) = \dfrac{1}{r}\dfrac{\sin\theta\cos\theta}{\sin^4\theta+\cos^4\theta}$. Since g is not integrable over $[0,\infty)\times[0,2\pi)$ (because r^{-1} is not integrable over $[0,\infty)$), f is not integrable over $\mathbb{R}^2$.

Example 8.12. Let $f(x,y) = \dfrac{x^2-y^2}{(x^2+y^2)^2}$. The square $(0,1)\times(0,1)$ is not very convenient for polar coordinates, but we can easily overcome this problem. Since f is bounded, hence integrable, over the bounded region $\{(x,y) : 0 < x < 1, 0 < y < 1, 1 < x^2+y^2\}$, f is integrable over the square if and only if it is integrable over the quadrant $E = \{(x,y) : 0 < x < 1, 0 < y < 1, x^2+y^2 \le 1\}$. In the notation of Theorem 8.9, $E' = \{(r,\theta) : 0 < r \le 1, 0 < \theta < \pi/2\}$ and

$$g(r,\theta) = r\,\frac{r^2(\cos^2\theta-\sin^2\theta)}{r^4} = \frac{\cos 2\theta}{r}.$$

Since r^{-1} is not integrable over $(0,1)$, g is not integrable over the rectangle E' (in (r,θ)-space), so f is not integrable over E.

Now we state a version of Theorem 8.9 for general changes of coordinates. Let $T : (u,v) \mapsto (x,y)$ be a change of variables, and suppose that x, y are differentiable functions of u, v. Let J_T be the Jacobian matrix:

$$J_T = \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}.$$

Observe that $J_{S\circ T} = J_S J_T$ (Chain Rule).

Theorem 8.13. *Let E' be an open subset of $\mathbb{R}^2$, $T : E' \to \mathbb{R}^2$ be a one-to-one differentiable function of E' onto a subset E of $\mathbb{R}^2$, and $f : E \to \mathbb{R}$ be a function. Then f is integrable over E if and only if $(f\circ T)|\det J_T|$ is integrable over E'. In that case,*

$$\int_E f = \int_{E'} (f\circ T)|\det J_T|.$$

Writing $\frac{\partial(x,y)}{\partial(u,v)}$ for $\det J_T$, this formula becomes

$$\int_E f(x,y)\,d(x,y) = \int_{E'} f(u,v)\left|\frac{\partial(x,y)}{\partial(u,v)}\right|\,d(u,v).$$

To recover Theorem 8.9 from Theorem 8.13, take $T(r,\theta) = (r\cos\theta, r\sin\theta)$, so $\frac{\partial(x,y)}{\partial(r,\theta)} = r$.

In the situation of Theorem 8.13, E is always measurable (continuous image of a Borel set) although this is not obvious.

One can extend Section 8 to $\mathbb{R}^n$ instead of $\mathbb{R}^2$. Moreover, for any (σ-finite) measure spaces $(\Omega_1, \mathcal{F}_1, \mu_1)$ and $(\Omega_2, \mathcal{F}_2, \mu_2)$, one can define a product $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \times \mu_2)$ such that Fubini's and Tonelli's theorems hold.

9. L^p-SPACES

A useful measure of distance between two integrable functions f and g is:

$$d(f,g) = \int |f-g| =: \|f-g\|_1.$$

Then

(i) $\|f\|_1 = 0$ if and only if $f = 0$ a.e. (Proposition 5.1(5),(7));
(ii) $\|\alpha f\|_1 = |\alpha| \|f\|_1$;
(iii) $\|f+g\|_1 \le \|f\|_1 + \|g\|_1$.

Consequently,

(i)′ $d_1(f,g) = 0$ if and only if $f = g$ a.e.
(ii)′ $d_1(g,f) = d_1(f,g)$;
(iii)′ $d_1(f,h) \le d_1(f,g) + d_1(g,h)$.

So $\|\cdot\|_1$ is almost a norm and d_1 is almost a *metric* (cf., Metric Spaces). The problems are that we have not yet defined a suitable vector space, and $\|f\|_1 = 0$ does not imply that f is the zero function.

If we allow our integrable functions to take the values ∞ and $-\infty$, then $f+g$ may not be everywhere defined (but it is almost everywhere defined). Any integrable function is real-valued almost everywhere, so we will now take $\mathcal{L}^1$ to be the space of all integrable functions with real (or complex) values. Then we identify functions which are almost everywhere equal (actually, we have effectively been doing this for some time). Define an equivalence relation on $\mathcal{L}^1$ by

$$f \sim g \iff f = g \text{ a.e.}$$

Let $[f]$ be the equivalence class of f, and $\mathcal{N} = [0] = \{f : \mathbb{R} \to \mathbb{R} : f = 0 \text{ a.e.}\}$. Then $\mathcal{N}$ is a subspace of the vector space $\mathcal{L}^1$, and we can form the quotient space $L^1 := \mathcal{L}^1/\mathcal{N}$ as a vector space whose elements are the equivalence classes $[f]$ (cf., Linear Algebra). Let

$$\|[f]\|_1 = \int |f|.$$

Then $\|\cdot\|_1$ is well-defined, and it is a norm on L^1. The distinction between $[f]$ and f is usually a distracting nuisance, so we suppress it, and we just write $\|f\|_1$ as the norm of f. However it is occasionally necessary to be aware of the difference.

Now we have a notion of convergence:

$$f_n \to f \text{ in } L^1\text{-norm} \iff \lim_{n\to\infty} \|f_n - f\|_1 = 0 \iff \int |f_n - f| \to 0.$$

In probability this may be called *convergence in mean*. Actually, convergence in mean square is more convenient in some respects. For that, one considers the space $\mathcal{L}^2$ of all measurable functions f such that $|f|^2$ is integrable. Suppose that $f, g \in \mathcal{L}^2$. Then simple inequalities for real/complex numbers give

$$|f+g|^2 \leq 2(|f|^2 + |g|^2), \qquad |f\overline{g}| \leq \tfrac{1}{2}(|f|^2 + |g|^2)$$

So $f + g \in \mathcal{L}^2$ and $f\overline{g}$ is integrable. Thus $\mathcal{L}^2$ is a vector space, and we can put

$$\langle f, g \rangle_2 = \int f\overline{g}.$$

Then $\langle \cdot, \cdot \rangle_2$ is positive-semidefinite, linear in the first variable, and conjugate-symmetric, so it is almost an inner product. Again there is a small problem that $\langle f, f \rangle_2 = 0$ implies only that $f \in \mathcal{N}$. So we form $L^2 = \mathcal{L}^2/\mathcal{N}$, and we obtain an inner product on L^2. Hence, we get a well-defined norm on L^2 given by

$$\|[f]\|_2 = \|f\|_2 = \langle f, f \rangle_2^{1/2} = \left(\int |f|^2 \right)^{1/2}.$$

Now, $\|f_n - f\|_2 \to 0$ (convergence in L^2-norm) corresponds exactly to convergence in mean square in the case of probability spaces.

Let's see what happens if the indices 1 and 2 are replaced by some other real $p > 0$. Let $\mathcal{L}^p$ be the set of all measurable functions f such that $|f|^p$ is integrable. Note that

$$(|f+g|)^p \leq (2\max(|f|, |g|))^p = 2^p \max(|f|^p, |g|^p) \leq 2^p(|f|^p + |g|^p),$$

$\mathcal{L}^p$ is a vector space. Let $L^p = \mathcal{L}^p/\mathcal{N}$, and

$$\|f\|_p = \left(\int |f|^p \right)^{1/p}.$$

Now it is not obvious whether the triangle inequality holds.

Proposition 9.1. [Minkowski's Inequality] *For $p \geq 1$ and $f, g \in \mathcal{L}^p$, $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.*

Proof. If $f = 0$ a.e. or $g = 0$ a.e., the inequality is trivial. So suppose that $\alpha := \|f\|_p > 0$ and $\beta := \|g\|_p > 0$.

The function $t \mapsto t^p$ is continuous on $[0, \infty)$ and its second derivative $p(p-1)t^{p-2}$ is positive on $(0, \infty)$. This implies that it is convex, i.e.

$$(\lambda s + (1-\lambda)t)^p \leq \lambda s^p + (1-\lambda)t^p$$

for $0 \leq \lambda \leq 1$, $s, t \geq 0$. Apply this with

$$\lambda = \frac{\alpha}{\alpha+\beta}, \quad s = \frac{|f(x)|}{\alpha}, \quad t = \frac{|g(x)|}{\beta}.$$

This gives

$$\left(\frac{|f| + |g|}{\alpha+\beta} \right)^p \leq \frac{1}{\alpha+\beta} \left(\frac{|f|^p}{\alpha^{p-1}} + \frac{|g|^p}{\beta^{p-1}} \right).$$

Using $|f+g| \leq |f| + |g|$, integrating, and taking p^{th} roots gives the required inequality. □

So L^p becomes a normed vector space, whenever $p \geq 1$.

We can also define a normed space for $p = \infty$. Let

$$\mathcal{L}^\infty(E) = \{f : E \to \mathbb{R} : f \text{ measurable and there exists } C > 0 \text{ such that } |f| \leq C \text{ a.e.}\}$$

This consists of the *essentially bounded* measurable functions on E. Given $f \in \mathcal{L}^\infty(E)$, define

$$\|f\|_\infty = \inf\{C > 0 : |f| \leq C \text{ a.e.}\}.$$

This quantity is known as the *essential supreumum* of $|f|$ (sometimes denoted $\text{esssup}|f|$) and for $f \in \mathcal{L}^\infty(E)$, $\|f\|_\infty = 0$ if and only if $f = 0$ a.e. Then, just as for finite p, we can define $L^\infty(E) = \mathcal{L}^\infty(E)$, and then $\|[f]\|_\infty = \|f\|_\infty$ gives a well defined norm on $L^\infty(E)$ (exercise).

Another important inequality, proved in a similar spirit to Minkowski's inequality is:

Proposition 9.2. [Hölder's Inequality] *Let $p, q \in (1, \infty)$ with $1/p + 1/q = 1$. Let $f \in L^p$ and $g \in L^q$. Then $fg \in L^1$ and $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

The pair (p, q) are sometimes called *Hölder conjugates*. For $p = q = 2$, Hölder's Inequality is the Cauchy-Schwarz Inequality. Notice also that Hölder's inequality holds for the pair $p = 1$ and $q = \infty$.

Proof. Note first that the function $t \mapsto \log t$ is concave on $[0, \infty)$, because its second derivative $-t^{-2}$ is negative. Hence

$$\frac{1}{p} \log s + \frac{1}{q} \log t \leq \log \left(\frac{s}{p} + \frac{t}{q} \right).$$

Exponentiate to obtain $s^{1/p} t^{1/q} \leq \frac{s}{p} + \frac{t}{q}$. Let $s = (|f(x)|/\|f\|_p)^p$ and $t = (|g(x)|/\|g\|_q)^q$. This gives

$$\frac{|fg|}{\|f\|_p \|g\|_q} \leq \frac{|f|^p}{p\|f\|_p^p} + \frac{|g|^q}{q\|g\|_q^q}.$$

Integrate. □

Corollary 9.3. *If $1 \leq p_1 < p_2 < \infty$ and $f \in L^{p_2}(a, b)$, then $f \in L^{p_1}(a, b)$ and*

$$\|f\|_{p_1} \leq (b - a)^{\frac{1}{p_1} - \frac{1}{p_2}} \|f\|_{p_2}.$$

Hence if $f_n \in L^{p_2}(a, b)$ and $\|f_n\|_{p_2} \to 0$, then $\|f_n\|_{p_1} \to 0$.

Proof. Apply Proposition 9.2 to the functions $|f|^{p_1}$ and $\chi_{(a,b)}$, with $p = p_2/p_1$. Then raise both sides to the power $(1/p_1)$. □

The inclusion $L^{p_2}(a, b) \subset L^{p_1}(a, b)$ in Corollary 9.3 is strict: consider x^α on $(0, 1)$.

Corollary 9.3 holds if (a, b) is replaced by any *finite* measure space. However, $L^{p_1}(1, \infty)$ is not contained in $L^{p_2}(1, \infty)$ (exercise).

For $p \geq 1$, L^p is a normed space and hence a metric space for $d_p(f, g) = \|f - g\|_p$. How does convergence in L^p-norm compare with pointwise a.e. convergence?

Examples 9.4. 1. Convergence a.e. does not imply convergence in L^p-norm: If $f_n(x) = n^2x^n(1-x)$ $(0 \le x \le 1)$, then $f_n(x) \to 0$ a.e., but $\|f_n\|_1 \to 1$.
2. Convergence in L^p-norm does not imply convergence a.e.: For $n = 2^r + k$, where $0 \le k < 2^r$, let f_n be the characteristic function of $[k2^{-r}, (k+1)2^{-r}]$. Then $\|f_n\|_1 = 2^{-r} \le 2/n \to 0$, but for each $x \in [0,1]$, $f_n(x)$ takes the values 0 and 1 infinitely often.

Theorem 9.5. *Let $p \in [1,\infty)$, and let (f_n) be a sequence in $\mathcal{L}^p$ which is Cauchy, i.e., for each $\varepsilon > 0$, there exists N such that $\|f_n - f_m\|_p < \varepsilon$ whenever $m, n \ge N$. Then there exists $f \in \mathcal{L}^p$ such that*

1. *There is a subsequence (f_{n_k}) such that $\lim_{k\to\infty} f_{n_k}(x) = f(x)$ a.e.*
2. $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Thus L^p is a complete metric space.

Proof. [For $p = 1$.] By assumption, there exist $N_1 < N_2 < N_3 < \dots$ such that $\int |f_n - f_m| < 2^{-(r+1)}$ whenever $n, m \ge N_r$. In particular, $\int |f_{N_{r+1}} - f_{N_r}| < 2^{-(r+1)}$. Let $g_1 = f_{N_1}$ and $g_r = f_{N_r} - f_{N_{r-1}}$ for $r = 2, 3, \dots$, so $\int |g_r| < 2^{-r}$ for $r \ge 2$. By Lebesgue's Series Theorem 6.8, $\sum_{r=1}^{\infty} g_r$ converges a.e. to $f \in \mathcal{L}^1$. Now

$$f_{N_k} = \sum_{r=1}^{k} g_r \to f \quad \text{a.e.},$$

$$\|f - f_{N_k}\|_1 = \int \left| \sum_{r=k+1}^{\infty} g_r \right| \le \int \sum_{r=k+1}^{\infty} |g_r| = \sum_{r=k+1}^{\infty} \int |g_r| < 2^{-k} \to 0.$$

If a Cauchy sequence has a convergent subsequence, then the whole sequence is convergent. See Prelims proof that every Cauchy sequence in $\mathbb{R}$ is convergent.

For general p, the use of LST has to be replaced by Minkowski's inequality plus Fatou's Lemma. □

Corollary 9.6. 1. *If $\|f_n - f\|_p \to 0$, then there is a subsequence (f_{n_r}) which converges to f a.e.*
2. *If $\|f_n - f\|_p \to 0$ and $f_n \to g$ a.e., then $f = g$ a.e.*

The Convergence Theorems provide situations when a.e. convergence implies convergence in L^p-norm. Here is a general result in that direction with a weaker conclusion (see the bonus sheet for a proof),

Theorem 9.7. [Egorov's Theorem] *Suppose that $f_n \to f$ a.e. Let E be a measurable set with $m(E) < \infty$ and let $\varepsilon > 0$. Then there is a measurable subset F of E with $m(E \setminus F) < \varepsilon$ such that $f_n \to f$ uniformly on F. In particular, $\|f_n - f\|_{L^p(F)} \to 0$ for all $p \ge 1$.*

It is very useful to identify natural dense subsets of the L^p spaces. We can often establish results on dense subsets and extend them to all of L^p by density (just as you often use density of $\mathbb{Q}$ and density of $\mathbb{R} \setminus \mathbb{Q}$ in prelims analysis arguments).

Theorem 9.8. *Let $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$.*

1. *There is a sequence of step functions* ψ_n *such that* $\lim_{n\to\infty} \|f - \psi_n\|_p = 0$.
2. *There is a sequence* (g_n) *of continuous functions with compact support*[12] *such that* $\lim_{n\to\infty} \|f - g_n\|_p = 0$.

Part 1 of this result is closely related to Theorem 3.11, that measurable functions are pointwise (a.e.) limits of sequences of step functions. For a proof when $p = 1$, see Stein & Shakarchi, Theorem 2.4, p.71.

As an example of density in action, we will show that translation of a function is continuous in the L^p norm. For $f : \mathbb{R} \to \mathbb{R}$, and $h \in \mathbb{R}$ consider the translation $f_h(x) = f(x - h)$. As Lebesgue measure is translation invariant, it follows that $f \in L^p(\mathbb{R})$ if and only if $f_h \in L^p(\mathbb{R})$ for all $h \in \mathbb{R}$.

Proposition 9.9. *For* $1 \le p < \infty$ *and* $f \in L^p(\mathbb{R})$, $\lim_{h\to 0} \|f_h - f\|_p = 0$.

Proof. Given $\epsilon > 0$, use Theorem 9.8(2) to find g which is continuous and of compact support such that $\|f - g\|_p < \epsilon/3$. As g is continuous and of compact support it is uniformly continuous. From this one obtains that $\lim_{h\to 0} \|g_h - g\|_p^p = 0$. So one can find $\delta > 0$ such that $\|g_p - g\|_p < \epsilon/3$ whenever $0 < |h| < \delta$. Using Minkowski's inequality and invariance of the Lebesgue measure under translation, for $0 < |h| < \delta$ one has

$$\|f - f_h\|_p \le \|f - g\|_p + \|g - g_h\|_p + \|g_h - f_h\|_p < 3\epsilon/3 = \epsilon. \quad \square$$

[You could equally use part (1) of Theorem 9.8 by showing that $\lim_{h\to 0} \|\psi_h - \psi\| = 0$ for a step function ψ (an easy calculation obtained by doing the integral when ψ is the indicator function of an interval, and then use a triangle inequality argument).]

We end this section by revisiting the Fourier transform of an integral function from the ASO Integral transforms course, giving rigorous proofs of some properties from that course. Let $f \in \mathcal{L}^1(\mathbb{R})$. The *Fourier transform* of f is the function $\widehat{f} : \mathbb{R} \to \mathbb{C}$ defined by

$$\widehat{f}(s) = \int_{\mathbb{R}} f(x) e^{-isx}\, dx.$$

Theorem 9.10. *Let* $f \in \mathcal{L}^1(\mathbb{R})$.

1. $|\widehat{f}(s)| \le \|f\|_1$ *for all* s,
2. $\widehat{f}$ *is continuous,*
3. $\widehat{f}(s) \to 0$ *as* $s \to \pm\infty$. **[Riemann–Lebesgue Lemma]**
4. *Let* $g(x) = xf(x)$. *If* $g \in \mathcal{L}^1(\mathbb{R})$ *then* $\widehat{f}$ *is differentiable everywhere and* $(\widehat{f})'(s) = -i\widehat{g}(s)$.
5. *If* f *has a continuous derivative* $f' \in \mathcal{L}^1(\mathbb{R})$, *then the Fourier transform of* f' *is* $is\widehat{f}(s)$.

Proof. (1) follows from $|f(x)e^{-isx}| = |f(x)|$. (2) follows from the continuous-parameter DCT (Theorem 7.2) with $g(x) = |f(x)|$.[13]

[12] i.e. the set $\overline{\{x \in \mathbb{R} : g(x) \neq 0\}}$ is compact.

[13] or alternatively by observing that $\widehat{f}$ is a uniform limit of continuous functions $\widehat{\varphi}_n$ where φ_n are step functions converging to f in L^1-norm.

For (3) when $f = \chi_{(a,b)}$, $\widehat{f}(s) = \dfrac{i(e^{-isb} - e^{-isa})}{s} \to 0$ as $|s| \to \infty$. This extends to step functions, by linearity. For general $f \in \mathcal{L}^1(\mathbb{R})$ and $\varepsilon > 0$, there is a step function φ such that $\|f - \varphi\|_1 < \varepsilon$ by Theorem 9.8, and there exists K such that $|\widehat{\varphi}(s)| < \varepsilon$ whenever $|s| > K$. Then

$$|\widehat{f}(s)| \le |\widehat{f}(s) - \widehat{\varphi}(s)| + |\widehat{\varphi}(s)| \le \|f - \varphi\|_1 + |\widehat{\varphi}(s)| < 2\varepsilon.$$

(4) can be proved by applying Theorem 7.5 with $|g|$ as dominating function. (5) can be proved by using integration by parts over intervals $[a_n, b_n]$ where $a_n \to -\infty$, $f(a_n) \to 0$, $b_n \to \infty$ and $f(b_n) \to 0$. □

One can alternatively prove (3) using the L^1-continuity of translations of Proposition 9.9 as follows. For $f \in L^1(\mathbb{R})$, making the change of variables $y = x + \pi/s$, we have

$$\hat{f}(s) = \int_{\mathbb{R}} f(x)e^{-isx}\, dx = \int_{\mathbb{R}} -f(x)e^{-isx-i\pi}\, dx = \int_{\mathbb{R}} -f(y - \pi/s)e^{-isy}\, dy.$$

Therefore Proposition 9.9 gives

$$|\hat{f}(s)| = \frac{1}{2}\left|\int_{\mathbb{R}} (f(x) - f(x - \pi/s))e^{-isx}\, dx\right| \le \frac{1}{2}\|f - f_{\pi/s}\|_1 \to 0,$$

as $s \to \infty$,

The theorem about the Fourier transform of the convolution of two integrable functions (Theorem 81) is an application of Fubini/Tonelli. One can also formulate a Fourier inversion theorem (normalising appropriately) when both f and $\hat{f}$ are integrable. See Stein and Shakarchi section 2.4. In fact Fourier inversion works particularly well in the L^2-setting; this will be further developped in the Fourier analysis course (and for Fourier series in the functional analysis course B4.2).

10. Absolutely continuous functions

This section consists of non-examinable material.

Recall from Section 4 that the Fundamental Theorem of Calculus is true for functions with a continuous derivative on $[a, b]$ (Theorem 4.11, but proved in Prelims), but it is false for the Cantor-Lebesgue function Φ whose derivative exists and equals 0 a.e. on $[0, 1]$ (Example 4.12).

The ideal Fundamental Theorem of Calculus would identify a class $\mathcal{A}$ of functions F on $[a, b]$ with both the following properties:

(i) If $F \in \mathcal{A}$, then F is differentiable a.e., $F' \in L^1(a, b)$, and $\int_a^x F'(y)\, dy = F(x) - F(a)$ for all $x \in [a, b]$.
(ii) If $f \in L^1(a, b)$ and $F(x) = \int_a^x f(y)\, dy$ for $x \in [a, b]$, then $F \in \mathcal{A}$ and $F' = f$ a.e.

It is not obvious that such a class exists—its existence implies that the indefinite integral F of an integrable function f is differentiable a.e. and $F' = f$ a.e.

In fact, this is true. Then $\mathcal{A}$ is the class of all functions of the form $F(x) := c + \int_a^x f$ for some $c \in \mathbb{R}$ and some $f \in L^1(a, b)$. Remarkably there is an intrinsic characterisation of such functions.

Let I be an interval. A function $F : I \to \mathbb{R}$ is said to be *absolutely continuous* on I if, for each $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sum_{r=1}^{n} |F(b_r) - F(a_r)| < \varepsilon$$

whenever $n \in \mathbb{N}$, (a_r, b_r) $(r = 1, \ldots, n)$ are disjoint subintervals of I and $\sum_{r=1}^{n}(b_r - a_r) < \delta$.

If we only allowed $n = 1$ in this definition, we would have the definition of uniform continuity on I. Recall from Prelims that any continuous function on $[a, b]$ is uniformly continuous.

Examples 10.1. 1. Recall that F is Lipschitz if there exists c such that $|F(y) - F(x)| \leq c|y - x|$ for all x, y. Any Lipschitz function is absolutely continuous (take $\delta = \varepsilon/c$).
2. If f is a bounded measurable function and $F(x) = \int_a^x f(y)\, dy$, then F is Lipschitz.
3. The Cantor-Lebesgue function is not absolutely continuous on $[0, 1]$.

Theorem 10.2. *Let $f \in L^1(I)$ and $F(x) = \int_a^x f(y)\, dy$. Then F is absolutely continuous on I.*

Theorem 10.3. *Let F be an absolutely continuous function on $[a, b]$. Then F is differentiable a.e., $F' \in L^1(a, b)$ and $F(x) - F(a) = \int_a^x F'(y)\, dy$ for all $x \in [a, b]$.*

One way to a proof of Theorem 10.2 is outlined in an optional exercise on the Supplementary Problem Sheet. There are various other proofs.

Theorem 10.3 is rather hard to prove. It is a remarkable theorem as differentiability (a.e.) is inferred from an assumption that seems to be only a type of continuity. Chapter 3 of Stein and Shakarchi goes into this in detail. I recommend this for further reading if you've enjoyed this course.

A corollary of Theorem 10.3 is that every Lipschitz function is differentiable a.e. Thus the Lipschitz functions are precisely the indefinite integrals of bounded measurable functions. (You may see a space of Lipchitz functions again in the functional analysis courses; you can use these ideas to identify the Lipschitz functions $f : [-1, 1] \to \mathbb{R}$ with $f(0) = 0$ and an appropriate norm with $L^\infty[-1, 1]$).