Lebesgue integral theory

Soonsik Kwon and Hyewon Yoon

This note is based on the lecture given by Soonsik Kwon. We use a book,

Frank Jones, *Lebesgue Integration on Euclidean space*, as a textbook.

In addition, other references on the subject were used, such as

- G. B. Folland, *Real analysis : mordern techniques and their application*, 2nd edition, Wiley.
- E. M. Stein, R. Shakarchi, *Real analysis*, Princeton Lecture Series in Analysis 3, Princeton Univercity Press.
- T. Tao, *An epsilon of room 1: pages from year three of a mathematical blog*, GSM **117**, AMS.
- T. Tao, *An introduction to measure theory*, GSM **126**, AMS.

CHAPTER 1

Introduction

1.1. Elements of Integrals

The main goal of this course is developing a 'new' integral theory. The development of the integral in most introductory analysis courses is centered almost exclusively on the Riemann integral. Riemann integral can be defined for some 'good' functions, for example, the spaces of functions which are continuous except finitely many points. However, in this course, we need to define an integral for a larger function class.

A typical integral consists of the following components:

The integrator

$$\int_A f(x)\, dx\,.$$

The set of integration (or the domain of integration)

The integrand

Roughly speaking, an integral is a summation of continuously changing object. Note the sign $\int$ represent the elongated S, the initial of 'sum'. It approximates

$$\sum_{\alpha \in I} \max_{x \in A_\alpha} f(x)|A_\alpha| \quad \text{or} \quad \sum_{\alpha \in I} \min_{x \in A_\alpha} f(x)|A_\alpha|,$$ [1]

where we decompose the set of integration A into disjoint sets, i.e., $A = \bigcup_{\alpha \in I} A_\alpha$ so that f is almost constant on each A_α. Here we denote $|A_\alpha|$ as the size of the set A_α. And $\max_{x \in A_\alpha} f(x)$ or $\min_{x \in A_\alpha} f(x)$ are chosen for the representatives of function value in A_α. Here we can naturally ask

QUESTION. How can we measure the size of a set?

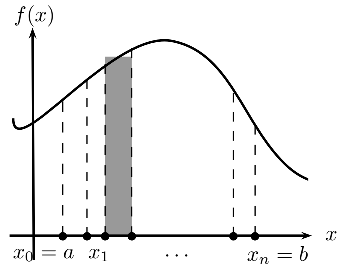
For the Riemann integral, we only need to measure the size of intervals (or rectangles, cubes for higher dimensions).

EXAMPLE 1.1. Let f be a $\mathbb{R}$-valued function from $[a, b]$ which is described in the next figure. First we chop out $[a, b]$ into small intervals $[x_0 = a, x_1]$, $[x_1, x_2]$, $\cdots$, $[x_{n-1}, x_n = b]$ and then we can approximate the value of integral by

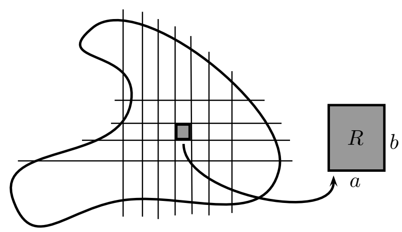
$$\sum_{i=1}^{n} f(x_i)|x_i - x_{i-1}|.$$

The limit of this sum will be defined to be the value of the integral and it will be called the Riemann integral. Here we use intervals to measure the size of sets in $\mathbb{R}$.

[1] one can replace $\max_{x \in A_\alpha}$ by any representative value of f in A_α



For the $\mathbb{R}^2$, we chop out the domain into small rectangles and use the area of the rectangles to measure the size of the sets.



For some 'ugry' functions (highly discontinuous functions), measuring size of intervals (or rectangles, cubes for higher dimension) is not enough to define the integral. So we want to define the size of the set for larger class of the good sets.

Then, on what class of subsets in $\mathbb{R}^n$ can we define the size of sets? From now on, we will call 'the size' of sets 'the measure' of sets. We want to define a *measure function*

$$m : \mathcal{M} \to [0, \infty],$$

where $\mathcal{M} \subseteq \mathcal{P}(\mathbb{R}^n)$, i.e., a subcollection of $\mathcal{P}(\mathbb{R}^n)$. Hence, our aim is finding a reasonable pair $(m, \mathcal{M})$ for our integration theory.

Can we define m for whole $\mathcal{P}(\mathbb{R}^n)$? If it is not possible, at least, we want to construct a measure on $\mathcal{M}$ appropriately so that $\mathcal{M}$ contains all of 'good' sets such as intervals (rectangles in higher dimension), open sets, compact sets. Furthermore, we hope the extended measure function m to agree with our intuition. To illustrate a few,

- $m(\varnothing) = 0$
- $m([a, b]) = b - a$, or $m(\text{Rectangle}) = |\text{vertical side}| \times |\text{horizontal side}|$
- If $A \subset B$, then $m(A) \leq m(B)$.
- If A, B are disjoint, then $m(A \cup B) = m(A) + m(B)$.

  Furthermore, we expect countable additivity:
- If $A_i, \quad i = 1, 2, \cdots$ are disjoint, then $\sum_{i=1}^{\infty} m(A_i) = m(\cup_{i=1}^{\infty} A_i)$.

In summary, in order to have a satisfactory integral theory we need to construct a measure function defined in a large class of subsets in $\mathbb{R}^n$. In Chapter 2, we construct the Lebesgue measure and proceed to its integral theory. Before then, we briefly review an 'old' theory, Riemann integrals.

1.2. A Quick review of Riemann integrals

Here we recall definitions and key theorems in Riemann integrals and observe some of its limitation. I ask you look back your textbook of Analysis course to recall proofs. Later we will revisit Riemann integrals in order to compare with Lebesgue integral after we develop Lebesgue theory. For simplicity, we will work on only $\mathbb{R}$. (The case for higher dimension will be very similar.) Suppose we have a bounded function $f : [a, b] \to \mathbb{R}$ such that $|f(x)| \le M$ for all $x \in [a, b]$.

Define a *partition* p of $[a, b]$ by $p = \{x_0, x_1, \cdots, x_n : a = x_0 < x_1 < \cdots < x_n = b\}$. We continue to define the *upper Riemann sum* with respect to a partition p by

$$\mathrm{URS}(f;p) := \sum_{i=1}^{n} \max_{x \in [x_{i-1}, x_i]} f(x)|x_i - x_{i-1}|.$$

The *lower Riemann sum* can be defined in a similar way by

$$\mathrm{LRS}(f;p) := \sum_{i=1}^{n} \min_{x \in [x_{i-1}, x_i]} f(x)|x_i - x_{i-1}|.$$

By the definition, we can easily see that

$$-M(b-a) \le \mathrm{LRS}(f;p) \le \mathrm{URS}(f;p) \le M(b-a)$$

for any partition p of $[a, b]$. We can also define a refinement partition p_* of p if $p_* \supset p$.

Example 1.2. Let p_1 and p_2 be partitions of $[a, b]$. Then $p_1 \cup p_2$ is also a partition of $[a, b]$ and moreover $p_1 \cup p_2$ is a refinement of p_1 and p_2. Note that the collection $\mathcal{P}$ of all partitions is partially ordered by inclusion.

Note that

$$\mathrm{LRS}(f;p) \le \mathrm{LRS}(f;p_*) \le \mathrm{URS}(f;p_*) \le \mathrm{URS}(f;p).$$

Observe that difference between URS and LRS is getting smaller as we refine a partition.

Now we may define the *upper Riemann integral* by

$$\overline{\int_a^b} f(x)dx := \inf_{p \in \mathcal{P}} \mathrm{URS}(p;f).$$

Similarly, we can define the *lower Riemann integral* by

$$\underline{\int_a^b} f(x)dx := \sup_{p \in \mathcal{P}} \mathrm{LRS}(p;f).$$

Then we have

$$\underline{\int_a^b} f(x)dx \le \overline{\int_a^b} f(x)dx$$

from definition. Finally, we say f is *Riemann integrable* if

$$\underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx.$$

and moreover we define

$$\int_a^b f(x)dx := \underline{\int_a^b} f(x)dx = \overline{\int_a^b} f(x)dx.$$

Theorem 1.1. *If $f : [a, b] \to \mathbb{R}$ is continuous then f is integrable on $[a, b]$.*

PROOF. First, note that f is uniformly continuous on $[a,b]$, i.e., for given $\epsilon > 0$, we can choose δ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$.

Let p be a partition of $[a,b]$ such that $|x_i - x_{i-1}| < \delta$. Then

$$\mathrm{URS}(f;p) - \mathrm{LRS}(f;p) = \sum_{i=1}^{n} \left[\max_{x \in [x_{i-1}, x_i]} f(x) - \min_{x \in [x_{i-1}, x_i]} f(x) \right] |x_i - x_{i-1}| < \sum_{i=1}^{n} \epsilon |x_i - x_{i-1}| \leq \epsilon(b-a).$$

By choosing ϵ sufficiently small, we can make $\mathrm{URS}(f;p) - \mathrm{LRS}(f;p)$ arbitrarily small. □

Corollary 1.2. *If $f : [a,b] \to \mathbb{R}$ is piecewise continuous, i.e., it is continuous except finitely many points, then f is integrable on $[a,b]$.*

PROOF. To prove this theorem, we slightly modify the above proof. Let $\{y_1, \cdots, y_n\}$ be the set of points of discontinuity. Even if we cannot shrink the size of the difference between $\max_{x \in I} f(x)$ and $\min_{x \in I} f(x)$, we can still shrink the size of intervals in our partition which contains y_1, $\cdots$, y_n. Since our function is bounded,[2] we can estimate the difference between URS and LRS on the intervals around discontinuity

$$\leq 2M \cdot (\text{the length of intervals around discontinuity}).$$

And we can choose the length of those intervals to be arbitrarily small. □

Corollary 1.3. *If $f : [a,b] \to \mathbb{R}$ is continuous except countably many points, then f is integrable on $[a,b]$.*

PROOF. For the case when there are infinitely many points of discontinuity, we may use the convergent infinite series $\sum_{i=1}^{\infty} (\epsilon 2^{-i})/M$. But you still have to be cautious about the fact that there are infinitely many points of discontinuity but you have only finitely many intervals in your partition. □

Exercise 1.1. *Write the proof of Theorem 0.1, Corollary 0.2 and Corollary 0.3 in detail.*

EXAMPLE 1.3 (A function which is not Riemann integrable). Define

$$f_{\mathrm{Dir}}(x) := \begin{cases} 1 & \text{if } x \in \mathbb{Q}, \\ 0 & \text{if } x \in \mathbb{Q}^c. \end{cases}$$

Then on any interval $I \subset \mathbb{R}$,

$$\max_{x \in I} f_{\mathrm{Dir}}(x) = 1 \quad \text{and} \quad \min_{x \in I} f_{\mathrm{Dir}}(x) = 0.$$

Hence f_{Dir} is not Riemann integrable. Also, note that f_{Dir} is nowhere continuous.

Theorem 1.4. *Let $f : [a,b] \to \mathbb{R}$ be bounded.*
If f is monotonic, then f is Riemann integrable.

[2]Note that every piecewise continuous functions on the compact domain is bounded.

PROOF. We may assume that f is nondecreasing. Let $p = \{x_0, x_1, \cdots, x_n\}$ be a partition such that $|x_i - x_{i-1}| = \frac{b-a}{n}, \quad i = 1, 2, \cdots, n.$

$$\begin{aligned}\mathrm{URS}(f;p) - \mathrm{LRS}(f;p) &= \sum_{i=1}^{n}\left[\max_{x\in[x_{i-1},x_i]} f(x) - \min_{x\in[x_{i-1},x_i]} f(x)\right]|x_i - x_{i-1}| = \sum_{i=1}^{n} f(x_i) - f(x_{i-1})\frac{b-a}{n} \\ &= (f(b) - f(a))\frac{b-a}{n} \le 2M(b-a)/n \le \epsilon,\end{aligned}$$

by choosing n sufficiently large. □

The last thing we want to mention is

Theorem 1.5 (Fundamental theorem of calculus)**.** *(1) Let f be a continuous $\mathbb{R}$-valued function on $[a, b]$. Then*

$$F(x) := \int_a^x f(y)dy$$

is differentiable on (a, b) and moreover $F' = f$.[3]

(2) Let f be a Riemann integrable $\mathbb{R}$-valued function on $[a, b]$. Then

$$F(x) := \int_a^x f(y)dy$$

is continuous on $[a, b]$. Furthermore, if f is continuous at $x_0 \in (a, b)$, then F is differentiable at x_0 and $F'(x_0) = f(x_0)$.

PROOF. Exercise. □

REMARK 1.4 (Limitation of Riemann integrals). First of all, to be Riemann integrable, in 'most' of small intervals $\max f - \min f$ must be small enough. So, we can say that the Riemann integrability depends on the continuity of functions. In fact, f is Riemann integrable if and only if f is continuous *'almost everywhere'*, where the term *almost everywhere* to be defined later.

Second, in Riemann integration theory, we only consider only intervals (rectangles or cubes for higher dimensions) to decompose the domain of integration. So we needed to know how to measure intervals. But for some functions, other type of decomposition would be natural. For example, if we can define measures of the sets like $\mathbb{Q}$ or $\mathbb{Q}^c$, then we can naturally define

$$\int_0^1 f_{Dir}(x)dx := 1 \cdot |\mathbb{Q} \cap [0, 1]| + 0 \cdot |\mathbb{Q}^c \cap [0, 1]|.$$

Finally, if we define a sequence $\{f_n : [0, 1] \to \mathbb{R}\}$ of functions by

$$f_n(x) := \begin{cases} 1 & \text{if } x = p/q \text{ where } p,\, q \in \mathbb{Z} \text{ and } q \le n, \\ 0 & \text{otherwise.} \end{cases}$$

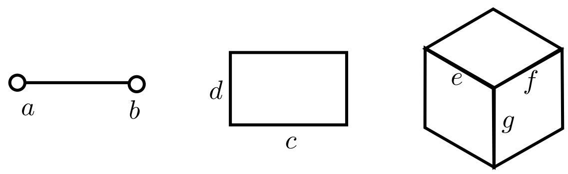
Then f_n's are Riemann integrable since they are continuous except finitely many points. Moreover $f_n \to f_{\mathrm{Dir}}$ as $n \to \infty$. But we have already seen that f_{Dir} is not Riemann integrable. This example shows us that even though all functions in the sequence is Riemann integrable, their limit can fail to be Riemann integrable. Note that if a sequences of Riemann integrable functions $\{f_n\}$ converge uniformly to f, then f is also Riemann integrable. (Exercise)

1.3. Jordan measure

We begin to construct a measure which agree with our intuition, such as length, area, or volume for each dimension. First we consider very restricted class of sets, so called elementary sets(or special polygons in *Jones*'s book).

The λ will be constructed to satisfy the following properties.

- $\lambda((a,b)) = \lambda((a,b]) = \lambda([a,b)) = \lambda([a,b]) = b - a =$ (the length of the interval),
- $\lambda(R) = cd =$ (the area of the rectangle),
- $\lambda(C) = efg =$ (the volume of the cube).

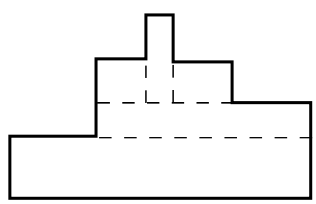


$I = (a,b) \in \mathbb{R}$ $\quad$ $R \in \mathbb{R}^2$ $\quad$ $C \in \mathbb{R}^3$

Definition 1.1. We say an *interval* is a subset of $\mathbb{R}$ of the form $[a,b], (a,b], [a,b)$, or (a,b). We define a measure of intervals by $\lambda(I) = b - a$. In higher dimensions, we define a *box* by a subset of $\mathbb{R}^n$ of the form $B = I_1 \times I_2 \times \cdots \times I_n$ where I_j are intervals. Then, we define a measure of boxes by

$$\lambda(B) = \prod_{j=1}^{n} \lambda(I_j).$$

An elementary set is any subset of $\mathbb{R}^n$ which is written as a finite union of boxes. [4]



P : an elementary set(or special polygon)

Then, the set of elementary sets in $\mathbb{R}^n$, denoted by $\mathcal{E}$, form a Boolean algebra. In other words, if $E, F \subset \mathbb{R}^n$ are elementary sets, then $E \cup F$, $E \cap F$, and $E \setminus F (= \{x \in \mathbb{R}^n : x \in E \text{ and } x \notin F\})$ are also elementary.(Exercise)

In order to define a measure on a elementary set, we need the following:

Lemma 1.6. *Let $E \subset \mathbb{R}^n$ be an elementary set.*

(1) *E can be expressed as the finite union of disjoint boxes.*

[4] In Jones's book, boxes and elementary sets are referred as special rectangles and special polygons, respectively. In their definition, they consider only closed boxes and closed elementary set. Then, in many statements in the following, one has to change 'disjoint' to 'nonoverlapping' which means that two sets has disjoint interior.

(2) *If E is partitioned as a finite union $B_1 \cup B_2 \cup \cdots \cup B_k$ of disjoint boxes, then then the quantity $\sum_{i=1}^{k} \lambda(B_i)$ is independent of the partition. We define the measure of E by $\lambda(E) = \sum_{i=1}^{k} \lambda(B_i)$.*

PROOF. (1) We first prove for 1 dimensional case. Given any finite collection of intervals $I_1, \cdots, I_k$, one can place the $2k$ end points in increasing order. We see that there exists a finite collection of disjoint intervals $J_1, \cdots, J_{k'}$ such that each of $I_1, \cdots, I_k$ are a union of a sub collection of $J_1, \cdots, J_{k'}$. For higher dimension, we express E as the union $B_1, \cdots, B_k$ of boxes $B_i = I_{i1} \times \cdots \times I_{in}$. For each $j = 1, \cdots, n$, we use the one dimensional argument to express $I_{1j}, \cdots, I_{kj}$ as the union of sub collections of $J_{1j} \cdots, J_{k'_j j}$ of disjoint intervals. One can express $B_1, \cdots, B_k$ as finite unions of boxes $J_{i_1 1} \times \cdots, \times J_{i_n n}$, where $1 \le i_j \le k'_j$ for all $1 \le j \le n$.
(2) We use discretization argument. Observe that for any interval I, the length of I can be recovered by the limiting formula

$$\lambda(I) = \lim_{N\to\infty} \frac{1}{N} \#(I \cap \frac{1}{N}\mathbb{Z}),$$

where $\frac{1}{N}\mathbb{Z} = \{n/N : n \in \mathbb{Z}\}$. In higher dimension, for any box B, we see that

$$\lambda(B) = \lim_{N\to\infty} \frac{1}{N^n} \#(B \cap \frac{1}{N}\mathbb{Z}^n).$$

If E is a finite disjoint union of $B_1, \cdots, B_k$, then

$$\lambda(E) = \sum_{j=1}^{k} \lim_{N\to\infty} \frac{1}{N^n} \#(B_j \cap \frac{1}{N}\mathbb{Z}^d) = \lim_{N\to\infty} \frac{1}{N^n} \#(E \cap \frac{1}{N}\mathbb{Z}^d).$$

In particular, $\lambda(E)$ is independent of decompositions of disjoint boxes. □

One can easily check fundamental properties: Let $E, F, E_j, j = 1, 2, \cdots, k$ be elementary sets.

- $\lambda(\varnothing) = 0$
- $\lambda(\{p\}) = 0$
- **P1**(monotonicity)

$$E \subset F \quad \Rightarrow \quad \lambda(E) \le \lambda(F)$$

- (finite subadditivity)

$$\lambda(E_1 \cup \cdots \cup E_k) \le \lambda(E_1) + \cdots + \lambda(E_k)$$

- **P2**(finite additivity) If E_j, $j = 1, \cdots, k$ are disjoint, then

$$\lambda(E_1 \cup \cdots \cup E_k) = \lambda(E_1) + \cdots + \lambda(E_k).$$

- (translation invariance)

$$\lambda(E + x) = \lambda(E)$$

So far, we have defined a measure function

$$\lambda : \mathcal{E} \to [0, \infty],$$

which satisfies fundamental conditions. However, the set of elementary sets is too restricted. We want to extend the measure function to a larger class of sets.
First, we discuss the Jordan measure, which associated to Riemann integrals.

DEFINITION 1.2. Let $E \subset \mathbb{R}^n$ be a bounded set.

- *Jordan inner measure*

$$\lambda_{*(J)}(E) = \sup_{\substack{A \subset E \\ A \text{ elementary}}} \lambda(A)$$

- *Jordan outer measure*

$$\lambda^{*(J)}(E) = \inf_{\substack{E \subset B \\ B \text{ elementary}}} \lambda(B)$$

- If $\lambda_{*(J)}(E) = \lambda^{*(J)}(E)$, then we say that E is *Jordan measurable*, define $\lambda(E)$ by the common number. We denote by $\mathcal{J}$, the collection of all Jordan measurable sets.

Note that we consider only bounded set so that Jordan outer measure to be defined. There is a way to extend Jordan measurability to unbounded sets, as this is not our final destination, we will not pursue this direction. One can observe elementary properties:

Lemma 1.7. *Let E, F be Jordan measurable.*

- *$E \cup F$, $E \setminus F$, $E \cap F$ are Jordan measurable.*
- $\lambda(E \cup F) \leq \lambda(E) + \lambda(F)$
- *If E and F are disjoint, then $\lambda(E \cup F) = \lambda(E) + \lambda(F)$.*
- *If $E \subset F$, then $\lambda(E) \leq \lambda(F)$.*
- *$E + x$ is Jordan measurable and $\lambda(E + x) = \lambda(E)$.*

Exercise 1.2. *Let E be a bounded set.*

(1) *Show that E and the closure $\overline{E}$ have the same Jordan outer measure.*
(2) *Show that E and the interior E° have the same Jordan inner measure.*
(3) *Show that E is Jordan integrable if and only if the topological boundary ∂E of E has Jordan measure zero.*

EXAMPLE 1.5.

(1) $[0,1] \cap \mathbb{Q}$ and $[0,1] \cap \mathbb{Q}^c$ are not Jordan measurable as the topological boundary is $[0,1]$.
(2) There are open sets that are not Jordan measurable. Let denote $[0,1] \cap \mathbb{Q} = \{r_n : n = 1, 2, \cdots\}$. Consider an open set $E = \cup_{n=1}^{\infty}(r_n - \frac{\epsilon}{2^(n+2)}, r_n + \frac{\epsilon}{2^{n+2}})$. The Jordan inner measure $\lambda_{*(J)}(E) \leq \epsilon$ but The Jordan outer measure $\lambda^{*(J)}(E) = 1$. Indeed, if $\cup_{i=1}^N I_i \supset E$, then $[0,1] \setminus \cup_{i=1}^N I_i$, which is also a finite union of disjoint intervals, cannot contain any non degenerate intervals by construction. Thus, $\lambda([0,1] - \cup_{i=1}^N I_i) = 0$ and so $\lambda(\cup_{i=1}^N I_i) = 1$. Hence, $\lambda^{*(J)}(E) = 1$. Later, we will see that E is Lebesgue measurable. Then, by countable additivity, one can show $\lambda(E) \leq \epsilon$.

    Similarly, considering $[0,1] \setminus E$, one can show that there are compact set which are not Jordan measurable.
(3) The above examples show that a countable union or a countable intersection of Jordan measurable sets may not be Jordan measurable. For instance, consider a sequence of elementary sets $E_n = \{x \in [0,1] : x = \frac{q}{p} \quad \text{where} \quad p \leq n\}$ or E_n^c.

Theorem 1.8. *(A connection to Riemann integral)*
E is a Jordan measurable set in $[a,b]$ if and only if the indicator function 1_E is Riemann integrable.

PROOF. Assume E is Jordan measurable. Fix $\epsilon > 0$. We can find elementary sets A, B such that $A \subset E \subset B$ and $\lambda(B \setminus A) \leq \epsilon$.(Prove!) Obviously, $1_A \leq 1_E \leq 1_B$ and $1_A, 1_B$ are Riemann integrable. Make a partition p consisting of end points of A, B. As A, B are elementary, $URS(1_A, p) = LRS(1_A, p) = \lambda(A)$, $URS(1_B, p) = LRS(1_B, p) = \lambda(B)$. Then we have $URS(1_E, p) \leq URS(1_B, p)$, $LRS(1_E, p) \geq LRS(1_A, p)$ and so we conclude that $URS(1_E, p) - LRS(1_E, p) \leq \epsilon$.
Conversely, assume that 1_E is Riemann integrable. For a fixed $\epsilon > 0$, we can find a partition p such that $\sum_{i=1}^N \max_{x_{i-1} \leq x \leq x_i} 1_E(x) - \min_{x_{i-1} \leq x \leq x_i} 1_E(x)(x_i - x_{i-1}) \leq \epsilon$. Choose elementary sets $A \subset E \subset B$ so that

$$B = \cup_i [x_{i-1}, x_i], \qquad \max 1_E = 1 \text{ on } [x_{i-1}, x_i]$$
$$A = \cup_i (x_{i-1}, x_i), \qquad \min 1_E = 1 \text{ on } [x_{i-1}, x_i].$$

Then, we have $\lambda(B) - \lambda(A) \leq \epsilon$. $\square$

CHAPTER 2

Lebesgue measure on $\mathbb{R}^n$

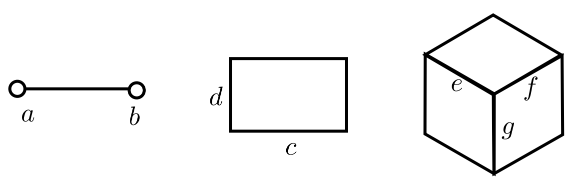
2.1. Construction

We want to extend 'measure' to a larger class of sets. We will denote the *Lebesgue measure*

$$\lambda : \mathcal{M} \to [0, \infty].$$

The λ will be constructed to satisfy the following properties.

- $\lambda((a,b)) = \lambda((a,b]) = \lambda([a,b)) = \lambda([a,b]) = b - a =$ (the length of the interval),
- $\lambda(R) = cd =$ (the area of the rectangle),
- $\lambda(C) = efg =$ (the volume of the cube).



$I = (a,b) \in \mathbb{R}$ $\quad$ $R \in \mathbb{R}^2$ $\quad$ $C \in \mathbb{R}^3$

We shall give the definition in six stages, progressing to more and more complicated classes of subsets of $\mathbb{R}^n$.

Stage 0 : The empty set. Define

$$\lambda(\varnothing) := 0.$$

Stage 1 : Special rectangles. In $\mathbb{R}^n$, a *special rectangle* is a closed cube of the form

$$I = [a_1, b_1] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n.$$

Note that each edge of a special rectangle is parallel to each axis. Define

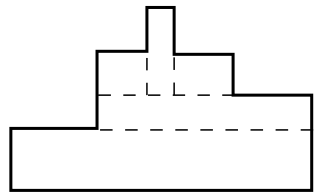
$$\lambda(I) := (b_1 - a_1) \cdot \cdots \cdot (b_n - a_n).$$

Stage 2 : Special Polygons. In $\mathbb{R}^n$, a *special polygon* is a finite union of nonoverlapping special rectangles. Here the word 'nonoverlapping' means having disjoint interiors, i.e., a special polygon P is the set of the form

$$P = \bigcup_{j=1}^{k} I_j,$$

where I_j's are nonoverlapping rectangles. Define

$$\lambda(P) := \sum_{j=1}^{k} \lambda(I_j).$$



P : a special polygon

One can naturally ask

QUESTION. Is $\lambda(P)$ well-defined?

For a given special polygon, there are several way of decomposition into special rectangles. Intuitively, it is an elementary but boring task to check the *well-definedness*. I leave it as an exercise. Furthermore, on the way to check it one can also show

Proposition 2.1 (**P1**, **P2**)**.** *Let P_1 and P_2 be special polygons such that $P_1 \subset P_2$. Then $\lambda(P_1) \leq \lambda(P_2)$. Moreover, if P_1 and P_2 are nonoverlapping each other, then $\lambda(P_1 \cup P_2) = \lambda(P_1) + \lambda(P_2)$.*

EXAMPLE 2.1. In $\mathbb{R}$, a special polygon is a finite union of nonoverlapping closed intervals. Write

$$P = \bigcup_{i=1}^{n} [a_i, b_i].$$

Then we can see that

$$\lambda(P) = \sum_{i=1}^{n} (b_i - a_i).$$

Exercise 2.1. *Prove the proposition 2.1.*

Stage 3 : Open sets. Let G be a nonempty open set in $\mathbb{R}^n$. Before we define Lebesgue measure on open sets, we observe the characterization of open sets. For one dimensional case, the structure of open sets is quite simple.

Proposition 2.2 (Problem 6 in the page 35 of the textbook)**.** *Every nonempty open subset G of $\mathbb{R}$ can be expressed as a countable disjoint union of open intervals.*

PROOF. For any $x \in G$, define $a_x := \inf\{a \in \mathbb{R} : (a, x) \subset G\}$ and $b_x := \sup\{b \in \mathbb{R} : (x, b) \subset G\}$. Here we allow a_x and b_x to be $\pm\infty$. Let $x \in I = (a, b) \subset G$. Then $I \subset I_x = (a_x, b_x)$. Indeed, $(a, x) \subset G$, $(x, b) \subset G$ and hence $a_x \leq a$ and $b \leq b_x$). Thus, $I \subset I_x$. So we can say that $I_x := (a_x, b_x)$ is the maximal interval in G containing x.

It is evident that $G \subset \bigcup_{x \in G} I_x$. On the other hand, for $y \in \bigcup_{x \in G} I_x$, there exists $z \in G$ such that $y \in I_z \subset G$. Thus

$$G = \bigcup_{x \in G} I_x.$$

Now we claim that $\bigcup_{x \in G} I_x$ is a countable disjoint union. First, assume that $I_x \cap I_y \neq \varnothing$ then there exists $z \in I_x \cap I_y$. Since I_z is the maximal interval in G containing z and $z \in I_x \subset G$, we have $I_x \subset I_z$. Similarly, we can see that $I_z \subset I_x$ and hence we get $I_z = I_x$. With the same argument, we can see that $I_z = I_y$. Therefore, $I_x = I_y$ or $I_x \cap I_y = \varnothing$, and so $\bigcup_{x \in G} I_x$ is a disjoint union. Also, by picking a rational number from each I_x, since $\mathbb{Q}$ is countable, we can conclude that G is a countable union of disjoint intervals. $\square$

Note that the decomposition above is unique.(Exercise)
In higher dimension, we have a weaker version of the above proposition.

Proposition 2.3. *Let $G \in \mathbb{R}^n$ be open. Then G is expressed by a countable union of non overlapping special rectangles.*

Proof. We use multi-index $(j) := (j_1, j_2, \cdots, j_n)$. We decompose $\mathbb{R}^n$ by special rectangles side of which is of length 2^{-k}. For $j_i \in \mathbb{Z}$, $k \in \mathbb{Z}_+$, denote

$$C_k^{(j)} := [\frac{j_1}{2^n}, \frac{j_1+1}{2^n}] \times \cdots \times [\frac{j_n}{2^n}, \frac{j_n+1}{2^n}].$$

For $k \in \mathbb{Z}_+$, we define inductively a index set $I_k = \{(j) : C_k^{(j)} \subset G, \text{ but } C_k^{(j)} \nsubseteq C_{k'}^{(j')} \text{ for any } (j') \in I_{k'}, k' < k\}$. Then, we claim that

$$G = \bigcup_{k=1}^{\infty} \bigcup_{(j) \in I_k} C_k^{(j)}.$$

For the proof, " $\supset$ " is obvious. The other inclusion is followed from openness of G. Indeed, for $x \in G$, there is a shrinking sequence of $\{C_k^{(j_k)}\}$ containing x. As there exist a ϵ−neighborhood of x, $B_\epsilon(x) \subset G$, one of $C_k^{(j_k)} \subset B_\epsilon(x) \subset G$. $\square$

$\lambda(G)$ will be obtained by approximating the measure of polygons within G. Define

$$\lambda(G) := \sup\{\lambda(P) : P \subset G, \ P \text{ is a special polygon}\}.$$

Note that there exists at least one special polygon $P \subset G$ with $\lambda(P) > 0$, since G is nonempty. So, $\lambda(G) > 0$ for any nonempty open set G. Also, even though $\lambda(P) < \infty$ for every $P \subset G$, $\lambda(G)$ could be ∞. For example, we have

$$\begin{aligned}
\lambda(\mathbb{R}^n) &= \sup\{\lambda(P) : P \subset \mathbb{R}^n\} \\
&\geq \sup\{\lambda([-a_1, a_1] \times \cdots [-a_n, a_n]) : a_1, \cdots, a_n \in \mathbb{R}\} \\
&= \sup\left\{2^n \prod_{i=1}^{n} a_i : a_1, \cdots, a_n \in \mathbb{R}\right\}.
\end{aligned}$$

Since $a_i > 0$ can be arbitrarily chosen, $\lambda(\mathbb{R}^n) = \infty$.

Here is the list of properties that λ satisfies.

Proposition 2.4. *Let G and G_k, $k = 1, 2, \cdots$, be open sets and P be a special poligon. Then the followings hold:*

*(**O1**) $0 \leq \lambda(G) \leq \infty$.*

*(**O2**) $\lambda(G) = 0$ if and only if $G = \varnothing$.*

*(**O3**) $\lambda(\mathbb{R}^n) = \infty$.*

*(**O4**) If $G_1 \subset G_2$, then $\lambda(G_1) \leq \lambda(G_2)$.*

*(**O5**)* $\lambda\left(\bigcup_{k=1}^{\infty} G_k\right) \leq \sum_{k=1}^{\infty} \lambda(G_k)$.

*(**O6**) If G_k's are disjoint, then* $\lambda\left(\bigcup_{k=1}^{\infty} G_k\right) = \sum_{k=1}^{\infty} \lambda(G_k)$.

*(**O7**)* $\lambda(P) = \lambda(P^\circ)$.

Proof. For **O4**, fix a special polygon $P \subset G_1$. Since P is also a special polygon in G_2, by definition, $\lambda(P) \leq \lambda(G_2)$. Hence $\lambda(G_1) = \sup\{\lambda(P) : P \subset G,\ P \text{ is a special polygon}\} \leq \lambda(G_2)$.

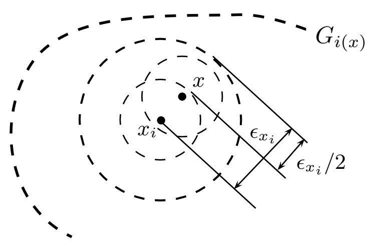
For **O5**, note that $\bigcup_{k=1}^{\infty} G_k$ is open and hence $\lambda\left(\bigcup_{k=1}^{\infty} G_k\right)$ can be defined. Fix a special polygon $P \subset \bigcup_{k=1}^{\infty} G_k$. For each $x \in P \subset \bigcup_{k=1}^{\infty} G_k$, $x \in G_{i(x)}$ for some index $i(x)$. Moreover, we can find ϵ_x so that $B(x, \epsilon_x) \subset G_{i(x)}$. Note that

$$\{B\left(x, \epsilon_x/2\right) : x \in P\}$$

is an open covering of P. Since P is compact, there exists a finite subcovering

$$\{B\left(x_i, \epsilon_{x_i}/2\right) : x_i \in P,\ i = 1, \cdots, N\}.$$

Let $\epsilon := \min\{\epsilon_{x_i}/2 : i = 1, \cdots, N\}$. For given $x \in P$, $x \in B(x_i, \epsilon_{x_i}/2)$ for some i and $B(x, \epsilon) \subset B(x_i, \epsilon_{x_i}) \subset G_{i(x)} \subset \bigcup_{k=1}^{\infty} G_k$.[1]



Let $P = \bigcup_{j=1}^{M} I_j$, where I_j's are nonoverlapping rectangles. We may assume that each I_j has the diameter[2] less than ϵ. (We can divide I_j into small rectangles whose diameter is less than ϵ.) Let x_j be the center of I_j. Then each $I_j \subset B(x_j, \epsilon) \subset G_k$ for some k. Merge I_j's which belong to G_k to form a new special rectangle Q_k. Indeed, we can define

$$Q_k := (\text{the union of } I_j\text{'s such that } I_j \subset G_k \text{ but } I_j \not\subset G_1, \cdots, G_{k-1})$$

Then each I_j is contained in one of G_k and $P = \bigcup_{k=1}^{\infty} Q_k$. In fact, P is a finite union of Q_k's. Suppose $Q_k = \varnothing$ for every $k \geq K$. Then

$$\lambda(P) = \sum_{k=1}^{K} Q_k \leq \sum_{k=1}^{K} \lambda(G_k) \leq \sum_{k=1}^{\infty} \lambda(G_k).$$

Since P is chosen arbitrarily, by definition,

$$\lambda\left(\bigcup_{k=1}^{\infty} G_k\right) \leq \sum_{k=1}^{\infty} \lambda(G_k).$$

---

[1] Such ϵ is called the Lebesgue number and the existence of such Lebesgue number is referred as the *Lebesgue number lemma*.

[2] In general, the diameter of a subset of a metric space is the least upper bound of the distances between pairs of points in the subset.

For **O6**, it suffices to show that

$$\sum_{k=1}^{\infty} \lambda(G_k) \leq \lambda\left(\bigcup_{k=1}^{\infty} G_k\right).$$

Fix N and then fix special polygons $P_1, \cdots, P_N$ such that $P_k \subset G_k$. Since G_k's are disjoint, so are P_k's. Note that $\bigcup_{k=1}^{N} P_k$ is a special polygon in $\bigcup_{k=1}^{\infty} G_k$. So we have

$$\sum_{k=1}^{N} \lambda(P_k) = \lambda\left(\bigcup_{k=1}^{N} P_k\right) \leq \lambda\left(\bigcup_{k=1}^{\infty} G_k\right).$$

Since $P_1, \cdots, P_N$ can be chosen arbitrarily,

$$\sum_{k=1}^{N} \lambda\left(G_k\right) \leq \lambda\left(\bigcup_{k=1}^{\infty} G_k\right).$$

Since N is arbitrary, finally we have

$$\sum_{k=1}^{\infty} \lambda\left(G_k\right) \leq \lambda\left(\bigcup_{k=1}^{\infty} G_k\right).$$

For **O7**, let P be a special polygon and write $P = \bigcup_{j=1}^{N} I_j$, where I_j's are nonoverlapping rectangles. First of all, it is obvious that $\lambda(P^\circ) \leq \lambda\left(P\right)$. To prove the other direction, fix $\epsilon > 0$. Then we can find a rectangle I'_j such that $I'_j \subset I_j^\circ$ and $\lambda(I'_j) \geq \lambda\left(I_j\right) - \epsilon$. For example, if $I_j = [a_1^{(j)}, b_1^{(j)}] \times \cdots \times [a_n^{(j)}, b_n^{(j)}]$, then we may take $I_j = [a_1^{(j)} - \delta, b_1^{(j)} + \delta] \times \cdots \times [a_n^{(j)} - \delta, b_n^{(j)} + \delta]$, where $0 < \delta < \epsilon/(2n)$. Since $\bigcup_{j=1}^{N} I'_j \subset P^\circ$, we get

$$\lambda\left(P^\circ\right) \geq \sum_{j=1}^{N} \lambda\left(I'_j\right) \geq \sum_{j=1}^{N} (\lambda(I_k) - \epsilon = \lambda(P) - N\epsilon.$$

Since ϵ is arbitrary, finally we obtain

$$\lambda\left(P\right) \leq \lambda\left(P^\circ\right). \qquad \square$$

Stage 4 : Compact sets. Let $K \in \mathbb{R}^n$ be a compact set. The Heine-Borel theorem asserts that a subset in a metric space is compact if and only if it is closed and bounded. Define

$$\lambda(K) := \inf\{\lambda\left(G\right) : K \subset G,\ G \text{ is open}\}.$$

For a special polygon P, since it is also a compact set, we have two definitions of $\lambda\left(P\right)$, as a special polygon and a compact set. We need to check two definitions coincide. Denote $\lambda_{\text{new}}(P)$ (resp. $\lambda_{\text{old}}(P)$) as a Lebesgue measure of P when we view P as a compact set (resp. a special rectangle).

Proposition 2.5. *For any special rectangle P, $\lambda_{new}(P) = \lambda_{old}(P)$.*

PROOF. First, let G be an open set such that $P \subset G$. Then, by definition, $\lambda_{\text{old}}(P) \leq \lambda(G)$. So

$$\lambda_{\text{old}}(P) \leq \inf\{\lambda\left(G\right) : P \subset G,\ G \text{ is open}\} = \lambda_{\text{new}}(P).$$

For the other direction, write $G = \bigcup_{j=1}^{N} I_j$, where I_j's are nonoverlapping rectangles. Fix $\epsilon > 0$. Then we may choose a closed rectangle I'_j, which is little bigger than I_j, so that $I_j \subset I_j'^\circ$ and

$\lambda\left(I_j'\right) \leq \lambda\left(I_j\right) + \epsilon$. Then $P \subset \bigcup_{j=1}^{N} I_j'^{\circ}$ and we have

$$\lambda_{new}(P) \leq \lambda\left(\bigcup_{j=1}^{N} I_j'^{\circ}\right) \leq \sum_{j=1}^{N} \lambda\left(I_j'^{\circ}\right) < \sum_{j=1}^{N} \lambda\left(I_j\right) + N\epsilon = \lambda_{\text{old}}(P) + N\epsilon.$$

Since ϵ is arbitrary, we can conclude that $\lambda_{new}(P) \leq \lambda_{\text{old}}(P)$ and hence $\lambda_{new}(P) = \lambda_{\text{old}}(P)$. □

Here is the list of properties that λ satisfies.

Proposition 2.6. *Let K, K_1 and K_2 be compact sets in $\mathbb{R}^n$. Then the followings hold:*

*(**C1**) $0 \leq \lambda(K) < \infty$.* *(**C2**) If $K_1 \subset K_2$, then $\lambda(K_1) \leq \lambda(K_2)$.*

*(**C3**) $\lambda(K_1 \cup K_2) \leq \lambda(K_1) + \lambda(K_2)$.*

*(**C4**) If K_1 and K_2 are disjoint, then $\lambda(K_1 \cup K_2) = \lambda(K_1) + \lambda(K_2)$.*

Proof. For **C1**, note that the equality sign in the right side is dropped because K is bounded.

For **C2**, let G be an open set containing K_2. Then $K_1 \subset G$ and $\lambda(K_1) \leq \lambda(G)$. So $\lambda(K_1) \leq \inf\{\lambda(G) : K_2 \subset G,\ G \text{ is open}\} = \lambda(K_2)$.

For **C3**, let G_1 and G_2 be open sets containing K_1 and K_2 respectively. Then $K_1 \cup K_2 \subset G_1 \cup G_2$. Here $K_1 \cup K_2$ is a compact set and $G_1 \cup G_2$ is an open set. So we have $\lambda(K_1 \cup K_2) \leq \lambda(G_1 \cup G_2) \leq \lambda(G_1) + \lambda(G_2)$. Since G_1 and G_2 can be arbitrarily chosen, we obtain $\lambda(K_1 \cup K_2) \leq \lambda(K_1) + \lambda(K_2)$.

For **C4**, it is enough to show that $\lambda(K_1) + \lambda(K_2) \leq \lambda(K_1 \cup K_2)$. Suppose that an open set G such that $K_1 \cup K_2 \subset G$ is given. Since K_1 and K_2 are disjoint open set, we may assume that $\text{dist}(K_1, K_2) = \epsilon > 0$. Let $G_1 := G \cap N(K_1, \epsilon/2)$[3] and $G_2 := G \cap N(K_2, \epsilon/2)$. Then $K_1 \subset G_1$, $K_2 \subset G_2$. Also, we can know that G_1 and G_2 are disjoint. So we have

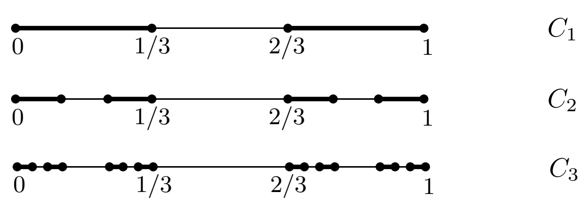
$$\lambda(K_1) + \lambda(K_2) \leq \lambda(G_1) + \lambda(G_2) = \lambda(G_1 \cup G_2) \leq \lambda(G).$$

Since G is an arbitrary open set containing $K_1 \cup K_2$, it follows that $\lambda(K_1) + \lambda(K_2) \leq \lambda(K_1 \cup K_2)$. and hence $\lambda(K_1) + \lambda(K_2) = \lambda(K_1 \cup K_2)$. □

Remark 2.2. Iterating the above proposition, we have $\lambda\left(\bigcup_{j=1}^{N} K_j\right) \leq \sum_{j=1}^{N} \lambda(K_j)$, if K_j's are compact sets. Eventually, we will have $\lambda\left(\bigcup_{j=1}^{\infty} K_j\right) \leq \sum_{j=1}^{\infty} \lambda(K_j)$. However, at this moment, $\bigcup_{j=1}^{\infty} K_j$ do not have to be compact and so $\lambda\left(\bigcup_{j=1}^{\infty} K_j\right)$ does not make sense.

Example 2.3 (Cantor ternary set). Let $G_1 = (1/3, 2/3)$, $G_2 = (1/3^2, 2/3^2) \cup (7/3^2, 8/3^2)$, $\cdots$ and define $C_1 = [0,1] - G_1$, $C_2 = C_1 - G_2$, $C_3 = C_2 - G_3$, $\cdots$. Then the *Cantor ternary set* is defined to be

$$C := \bigcap_{k=1}^{\infty} C_k = [0,1] - \bigcup_{k=1}^{\infty} G_k.$$



[3]Here $N(K_1, \epsilon/2)$ denotes an $\epsilon/2$-neighborhood of K_1, i.e., $N(K_1, \epsilon/2) := \{x \in \mathbb{R}^n : |x - y| < \epsilon/2 \text{ for all } k \in K_1\}$.

Thus C is compact. For each n, $C_n = [0,1] - \bigcup_{k=1}^{n} G_k$. So $\lambda(C) \le \lambda(C_n) = (2/3)^n$ for every n and hence $\lambda(C) = 0$.

Now observe the relation between the Cantor set and its ternary expansion. Every $x \in [0,1]$ can be written

$$x = \sum_{j=1}^{\infty} \frac{\alpha_j}{3^j},$$

where $\alpha_j = 0$, 1 or 2. We call this representation of x its *ternary expansion*. To simplify the notation, we express this equation symbolically in the form

$$x = 0.\alpha_1\alpha_2\alpha_3 \cdots .$$

The ternary expansion is unique except when a ternary expansion terminates, i.e., $\alpha_j = 0$ except finitely many indices j. For example,

$$\frac{1}{3} + \frac{1}{3^2} = \frac{1}{3} + \sum_{j=3}^{\infty} \frac{2}{3^j}.$$

Theorem 2.7. *Let $x \in [0,1]$. Then $x \in C$ if and only if x has a ternary expansion consisting only of 0's and 2's.*

PROOF. Write $x = \sum_{j=1}^{\infty} \frac{\alpha_j}{3^j}$. Observe that, for each k, if $x \in C_k$, then $\alpha_k \neq 1$. Therefore, if $x \in C$, then $\alpha_j \neq 1$ for every j.

Conversely, assume by the contradiction that $x \notin C$. Then $x \in \bigcup_{k=1}^{\infty} G_k$. In other words, $x \in G_k$ for some k. Now check the fact that

$$G_j = \left\{ x = \sum_{j=1}^{\infty} \frac{\alpha_j}{3^j} \in (0,1) : \alpha_j = 1 \text{ and } x \neq 0.\cdots\alpha_j 000\cdots,\ 0.\cdots\alpha_j 222\cdots \right\}.$$

Then it follows the contradiction. □

Proposition 2.8 (Problem 23 in the page 42 of the textbook)**.** *C is uncountable.*

PROOF. One of the ways to prove this claim is using a diagonal method in the ternary expansion. □

REMARK 2.4 (Hausdorff dimension of the Cantor set). The Cantor set stimulated a deeper study on geometric properties on sets. Indeed, one can generalize the notion of dimension to real numbers. It is called *Hausdorff dimension* and *Hausdorff measure*, which generalize α-dimensional Lebesgue measure. See, for instance, [**Fol**], [**StSh**], [**Tao**] for detail. For any $E \subset \mathbb{R}^n$, we define the *α-dimensional Hausdorff outermeasure* of E by

$$m_\alpha^*(E) := \liminf_{\delta \to 0} \left\{ \sum_k (\operatorname{diam} F_k)^\alpha : E \subset \bigcup_{k=1}^{\infty} F_k,\ \operatorname{diam} F_k \le \delta \text{ for all } k \right\},$$

which satisfies the countable additivity when one restricted to a measurable class.

In particular, if E is a closed set, it is known that there exists a unique α such that

$$m_\beta^*(E) = \begin{cases} \infty & \text{if } \beta < \alpha \\ 0 & \text{if } \alpha < \beta. \end{cases}$$ [4]

In this case, we say that E has *Hausdorff dimension* α. For instance, the Cantor set C has Hausdorff dimension $\log_3 2 < 1$.

Exercise 2.2. *Verify that the Cantor set C has Hausdorff dimension $\log_3 2 < 1$. Construct a set $E \subset \mathbb{R}^2$ having Hausdorff dimension $\log_3 5$, $\log_3 4$. For any $0 < r < \infty$, find a set having Haudorff measure r.*

We proceed to define the Lebesgue mesure,

Definition 2.1. Let $A \in \mathbb{R}$ be an arbitrary set. Define

$$\lambda^*(A) = \text{the } \textit{outer measure} \text{ of } A := \inf\{\lambda(G) : A \subset G,\ G \text{ is an open set}\}.$$
$$\lambda_*(A) = \text{the } \textit{inner measure} \text{ of } A := \sup\{\lambda(K) : K \subset A,\ K \text{ is a compact set}\}.$$

For any open set G and compact set K such that $K \subset A \subset G$, $\lambda(K) \le \lambda(G)$. It implies that $\lambda_*(A) \le \lambda^*(A)$. Let G be an open set and K be a compact set. Then

$$\begin{aligned} \lambda_*(G) = &\ \sup\{\lambda(K) : K \subset G,\ K \text{ is a compact set}\} \\ \ge &\ \sup\{\lambda(P) : P \subset G,\ P \text{ is a special polygon}\} = \lambda(G) = \lambda^*(G). \end{aligned}$$

So we have $\lambda^*(G) = \lambda(G) = \lambda_*(G)$. Also, we can obtain

$$\lambda^*(K) = \inf\{\lambda(G) : K \subset G,\ G \text{ is an open set}\} = \lambda(K) = \lambda_*(K).$$

Proposition 2.9. *Let A, A_k, $k = 1, \cdots,$ and B be subsets in $\mathbb{R}^n$. Then the followings hold:*

*(***2**) If $A \subset B$, then $\lambda^*(A) \le \lambda^*(B)$ and $\lambda_*(A) \le \lambda_*(B)$.*

*(***3**) $\lambda^*\left(\bigcup_{k=1}^\infty A_k\right) \le \sum_{k=1}^\infty \lambda^*(A_k)$.* *(***4**) If A_k's are disjoint, then $\lambda_*\left(\bigcup_{k=1}^\infty A_k\right) \ge \sum_{k=1}^\infty \lambda_*(A_k)$.*

Proof. For ***3**, fix $\epsilon > 0$. Then there exists open set G_k such that $A_k \subset G_k$ and $\lambda(G_k) < \lambda^*(A_k) + \epsilon 2^{-k}$. So we have

$$\lambda^*\left(\bigcup_{k=1}^\infty A_k\right) \le \lambda\left(\bigcup_{k=1}^\infty G_k\right) \le \sum_{k=1}^\infty \lambda(G_k) < \sum_{k=1}^\infty (\lambda^*(A_k) + \epsilon 2^{-k}) = \sum_{k=1}^\infty \lambda^*(A_k) + \epsilon.$$

Since ϵ can be chosen arbitrarily, we have $\lambda^*\left(\bigcup_{k=1}^\infty A_k\right) \le \sum_{k=1}^\infty \lambda^*(A_k)$.

For ***4**, fix N. For $j = 1, \cdots, N$, let $K_j \subset A_j$ be the compact sets. Then

$$\lambda_*\left(\bigcup_{k=1}^\infty A_k\right) \ge \lambda_*\left(\bigcup_{k=1}^N A_k\right) \ge \lambda\left(\bigcup_{k=1}^\infty K_k\right) = \sum_{k=1}^N \lambda(K_k).$$

Since K_j can be chosen arbitrarily, we have $\lambda_*\left(\bigcup_{k=1}^\infty A_k\right) \ge \sum_{k=1}^N \lambda_*(A_k)$. Also, since N is arbitrary, we can obtain $\lambda_*\left(\bigcup_{k=1}^\infty A_k\right) \ge \sum_{k=1}^\infty \lambda_*(A_k)$. $\square$

[4] In fact, there exists such unique α if E is a Borel set, i.e., E is contained in the smallest σ-algebra containing all open set. We will see the definition of σ-algebra soon or later.

Stage 5 : Sets having finite outer measure.

DEFINITION 2.2. Let $A \in \mathbb{R}^n$ with $\lambda^*(A) < \infty$. Define

$$\mathcal{L}_0 := \{A \in \mathbb{R}^n : \lambda_*(A) = \lambda^*(A) < \infty\},$$

that is, $\mathcal{L}_0$ is the class of measurable sets with finite measure. For $A \in \mathcal{L}_0$, define $\lambda(A) := \lambda_*(A) = \lambda^*(A)$.

Any open sets and compact sets with finite measure are contained in $\mathcal{L}_0$.

REMARK 2.5. Readers should understand why we have to confine to finite measure sets when we check measurability. Measurability is a local property. If we have a non-measurable set, by putting together countable translated copy of that, one can construct a set A with $\lambda^*(A) = \lambda_*(A) = \infty$. If not finite, the measurability condition is essentially void. We will see non-measurable sets such that $\lambda^*(A) = \lambda_*(A) = \infty$.

Lemma 2.10. *If A, $B \in \mathcal{L}_0$ are disjoint, then $A \cup B \in \mathcal{L}_0$ and $\lambda(A \cup B) = \lambda(A) + \lambda(B)$.*

PROOF. First, note that $\lambda^*(A \cup B) \geq \lambda_*(A \cup B)$. On the other hand,

$$\lambda^*(A \cup B) \leq \lambda^*(A) + \lambda^*(B) = \lambda(A) + \lambda(B) = \lambda_*(A) + \lambda_*(B) \leq \lambda_*(A \cup B).$$

So all the terms in the above must be equal. In particular, we have $A \cup B \in \mathcal{L}_0$ and $\lambda(A \cup B) = \lambda(A) + \lambda(B)$. $\square$

Theorem 2.11 (Approximation lemma for $\mathcal{L}_0$)**.** *Suppose that $A \in \mathbb{R}^n$ with $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$ if and only if there exist a compact set K and an open set G such that $\lambda(G \setminus K) < \epsilon$ for each $\epsilon > 0$.*

PROOF. First, assume that $A \in \mathcal{L}_0$. By definition of λ^* and λ_*, for each $\epsilon > 0$ there exist a compact set K and an open set G such that $\lambda(K) > \lambda_*(A) - \epsilon/2$ and $\lambda(G) < \lambda^*(A) + \epsilon/2$. Then $\lambda(G \setminus K) = \lambda(G) - \lambda(K) < \epsilon$.

Conversely, fix $\epsilon > 0$. Let K and G be a compact set and an open set such that $\lambda(G \setminus K) < \epsilon$ for each $\epsilon > 0$. Then we have

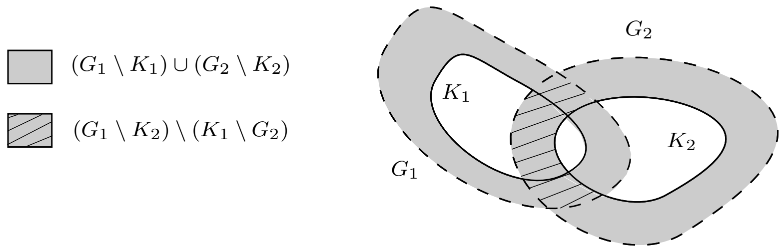
$$\lambda^*(A) \leq \lambda(G) = \lambda(K) + \lambda(G \setminus K) < \lambda(K) + \epsilon \leq \lambda_*(A) + \epsilon.$$

Since ϵ is arbitrary, we have $\lambda^*(A) \leq \lambda_*(A)$. $\square$

Corollary 2.12. *If A, $B \in \mathcal{L}_0$, then $A \cup B$, $A \cap B$, $A \setminus B \in \mathcal{L}_0$.*

PROOF. Fix $\epsilon > 0$. Let K_1 and G_1 (resp. K_2 and G_2) be a compact set and an open set such that $K_1 \subset A \subset G_1$ and $\lambda(G_1 \setminus K_1) < \epsilon/2$ (resp. $K_1 \subset A \subset G_2$ and $\lambda(G_2 \setminus K_2) < \epsilon/2$). Then $K_1 \setminus G_2$ is compact and $G_1 \setminus K_2$ is open. Moreover, $K_1 \setminus G_2 \subset A \setminus B \subset G_1 \setminus K_2$ and

$$(G_1 \setminus K_2) \setminus (K_1 \setminus G_2) \subset (G_1 \setminus K_1) \cup (G_2 \setminus K_2).$$



So we have

$$\lambda\left((G_1 \setminus K_2) \setminus (K_1 \setminus G_2)\right) \leq \lambda\left((G_1 \setminus K_1) \cup (G_2 \setminus K_2)\right) < \epsilon.$$

Therefore $A \setminus B \in \mathcal{L}_0$.

Also, by lemma 2.9, we can conclude that $A \cap B = A \setminus (A \setminus B) \in \mathcal{L}_0$ and $(A \setminus B) \cup B \in \mathcal{L}_0$. □

Theorem 2.13 (Countable subadditivity). *Suppose that $A_k \in \mathcal{L}_0$ for $k = 1, 2, \cdots$. Let $A := \bigcup_{k=1}^{\infty} A_k$, and assume $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$ and*

$$\lambda(A) \leq \sum_{k=1}^{\infty} \lambda(A_k).$$

In addition, if the A_k's are disjoint, then

$$\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k).$$

Proof. Assume that A_k's are disjoint. Then we have

$$\lambda^*(A) \leq \sum_{k=1}^{\infty} \lambda^*(A_k) \leq \sum_{k=1}^{\infty} \lambda_*(A_k) \leq \lambda_*(A) \leq \lambda^*(A).$$

So all the terms in the above must be equal. In particular, we get $\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k)$.

For the general case, let $B_1 := A_1, \cdots, B_k := A_k - (A_1 \cup \cdots \cup A_{k-1}), \cdots$. Then B_k's are disjoint sets in $\mathcal{L}_0$ and moreover $\bigcup_{k=1}^{\infty} B_k = A$. So we have

$$\lambda(A) = \sum_{k=1}^{\infty} \lambda(B_k) \leq \sum_{k=1}^{\infty} \lambda(A_k),$$

since $B_k \subset A_k$ for each k. □

Stage 6 : Arbitrary measurable sets.

Definition 2.3. Let $A \in \mathbb{R}^n$. We call A *measurable* if for all $M \in \mathcal{L}_0$, $A \cap M \in \mathcal{L}_0$. In case A is measurable, the *Lebesgue measure* of A is

$$\lambda(A) := \sup\{\lambda(A \cap M) : M \in \mathcal{L}_0\}.$$

Moreover, we denote $\mathcal{L}$ by the class of all measurable sets $A \in \mathbb{R}^n$.

Remark 2.6. One can later show by a property of $\mathcal{L}$ (**M2**) that $A \in \mathcal{L}$ if and only if $A \cap B_R \in \mathcal{L}_0$ for any ball B_R.

Of course, we have to check the consistency of this definition. In other words, we will show

Proposition 2.14. *Let $A \in \mathbb{R}^n$ with $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$ if and only if $A \in \mathcal{L}$. Moreover, the definition of $\lambda(A)$ in Stage 5 and 6 produce the same number.*

Proof. Suppose that $A \in \mathcal{L}_0$. For arbitrary $M \in \mathcal{L}_0$, $A \cap M \in \mathcal{L}_0$. Thus $A \in \mathcal{L}$. Conversely, assume that $A \in \mathcal{L}$. Let $B_k := B(0,k)$ for $k \in \mathbb{Z}_+$. Then, by definition, $A_k := A \cap B_k \in \mathcal{L}_0$ for each k. By countable additivity, $A = \bigcup_{k=1}^{\infty} A_k \in \mathcal{L}_0$.

To show the consistency of the definition of the measure, assume $A \in \mathcal{L}$ and let $\lambda(A)$ (resp. $\widetilde{\lambda}(A)$) stands for the measure of A which we have defined in Stage 5 (resp. Stage 6). Then we can see that

$$\widetilde{\lambda}(A) := \sup\{\lambda(A \cap M) : M \in \mathcal{L}_0\} \geq \lambda(A \cap A) = \lambda(A).$$

Also, since $A \cap M \subset A$ for each $M \in \mathcal{L}_0$, $\lambda(A \cap M) \leq \lambda(A)$ for each $M \in \mathcal{L}_0$ and we can conclude that $\widetilde{\lambda}(A) \leq \lambda(A)$. In conclusion, we have $\widetilde{\lambda}(A) = \lambda(A)$. □

2.2. Properties of Lebesgue measure

Proposition 2.15. *Let A, B and A_k, $k = 1, 2, \cdots$ be measurable sets($\in \mathcal{L}$) in $\mathbb{R}^n$. Then the followings hold:*

(**M1**) $A^c \in \mathcal{L}$.

(**M2**) $A := \bigcup_{k=1}^{\infty} A_k$, $B := \bigcap_{k=1}^{\infty} A_k \in \mathcal{L}$.

(**M3**) $A \setminus B \in \mathcal{L}$.

(**M4**) $\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \lambda(A_k)$.
If A_k's are disjoint, then $\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \lambda(A_k)$.

(**M5**) *If* $A_1 \subset A_2 \subset A_3 \subset \cdots$, *then* $\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \lambda(A_k)$.

(**M6**) *If* $A_1 \supset A_2 \supset \cdots$ *and* $\lambda(A_1) < \infty$ *,then* $\lambda\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \lambda(A_k)$.

Proof.
M1 Note that $A^c \cap M = M \setminus A = M \setminus (A \cap M) \in \mathcal{L}_0$ for any $M \in \mathcal{L}_0$.
M2 Let $M \in \mathcal{L}_0$ be given. Note that $A \cup M = \bigcup_{k=1}^{\infty}(A_k \cup M)$. Since $A_k \cap M \in \mathcal{L}_0$ for each k and $\lambda^*(A \cap M) \leq \lambda(M) < \infty$, Countable additivity of $\mathcal{L}_0$ implies that $A \cup M \in \mathcal{L}_0$. Since M is arbitrary, we can conclude that $A \in \mathcal{L}$. Proof for B is similar.
M3 Since $A \setminus B = A \cap B^c$, the statement **M3** directly follows form **M1** and **M2**.
M4 For given $M \in \mathcal{L}_0$, we have $\lambda\left(\bigcup_{k=1}^{\infty} A_k \cap M\right) \leq \sum_{k=1}^{\infty} \lambda(A_k \cap M) \leq \sum_{k=1}^{\infty} \lambda(A_k)$. Since $M \in \mathcal{L}_0$ is arbitrary, $\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \lambda(A_k)$.
If A_k's are disjoint, fix $N \in \mathbb{Z}_+$. Furthermore, fix $M_1, \cdots, M_N \in \mathcal{L}_0$. Denote $M = \bigcup_{k=1}^{N} M_k$. Then, we have $\lambda(A) \geq \lambda(A \cap M) = \sum_{k=1}^{N} \lambda(A_k \cap M) \geq \lambda\left(\bigcup_{k=1}^{N} A_k \cap M_k\right)$ from the countable additivity of $\mathcal{L}_0$. Since $M_1, \cdots, M_N$ are arbitrary, we have $\lambda(A) \geq \sum_{k=1}^{N} \lambda(A_k)$. Finally, as N is arbitrary, we conclude that $\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k)$.

M5 Express $\bigcup_{k=1}^{\infty} A_k$ as a disjoint union $A_1 \dot{\bigcup} \dot{\bigcup}_{k=2}^{\infty}(A_k \setminus A_{k-1})$. Then, apply **M4**:

$$\begin{aligned}\lambda\left(\bigcup_{k=1}^{\infty} A_k\right) &= \lambda(A_1) + \sum_{k=2}^{\infty} \lambda(A_k \setminus A_{k-1}) \\ &= \lim_{N\to\infty} \lambda\left(A_1 \dot{\bigcup} \dot{\bigcup}_{k=2}^{N} (A_k \setminus A_{k-1})\right) \\ &= \lim_{N\to\infty} \lambda(A_N)\end{aligned}$$

M6 Similar to the proof of $textbgM5$ Note that one has to use $\lambda(A_1) < \infty$. □

Proposition 2.16 (**M7**)**.** *All open sets and closed sets are contained in $\mathcal{L}$.*

Proof. Any open set G is a countable union of $G \cap B(0, i) \in \mathcal{L}_0$ for $i = 1, 2, \cdots$. Then use **M4**. A closed set is a complement of open set and so in $\mathcal{L}$ by **M2**.

□

Proposition 2.17 (**M8**)**.** *Let $A \in \mathbb{R}^n$. If $\lambda^*(A) = 0$, then $A \in \mathcal{L}$.*

Proposition 2.18 (Approximation property, **M9**)**.** *Let $A \in \mathbb{R}^n$. The followings are equivalent.*

(1) *A is measurable*
(2) *For every $\epsilon > 0$ there exists an open set G such that*

$$A \subset G \quad \text{and} \quad \lambda^*(G \setminus A) < \epsilon.$$

(3) *For every $\epsilon > 0$ there exists a closed set F such that*

$$F \subset A \quad \text{and} \quad \lambda^*(A \setminus F) < \epsilon.$$

Proof. (1)⇒(2)
Decompose A into $A_k = A \cap B(k, k-1)$ where $B(k, k-1) = \{x \in \mathbb{R}^n : k-1 \le |x| < k\}$. For each k, find open sets G_k such that $G_k \supset A_k$ with $\lambda(G_k \setminus A_k) \le \epsilon/2^k$. Then $G = \bigcup_{k=1}^{\infty} G_k$ is a desired open set satisfying $\lambda(G \setminus A) \le epsilon$.
(2)⇒ (1)
For each $k \in \mathbb{Z}_+$, find an open set $G_k \supset A$ such that $\lambda(G_k \setminus A) \le 1/k$. By **M6**, $\lambda(\bigcap_{k=1}^{\infty} G_k \setminus A) = 0$ and so $\bigcap_{k=1}^{\infty} G_k \setminus A \in \mathcal{L}$. Hence, $A \in \mathcal{L}$.
(1) ⇔ (3)
Use **M2** and the previous steps.

□

Remark 2.7. Indeed, in some other textbooks ([**StSh**], [**Tao**]), it is used for the definition of Lebesgue measurability. Note that For $A \in \mathcal{L}$, we can express $A = \bigcap_{k=1}^{\infty} G_k \cup N = \bigcup_{k=1}^{\infty} F_k \cup \tilde{N}$, where G_k's are open, F_k are closed and $N, \tilde{N}$ are measure zero sets.

Proposition 2.19 (**M10**)**.** *If $A \in \mathcal{L}$, then $\lambda(A) = \lambda_*(A) = \lambda^*(A)$.*

PROOF. We have already see that this statement is true when $A \in \mathcal{L}_0$. In case of $\lambda^*(A) = \infty$, suppose that $\lambda(A) = c < \infty$. Then, by **M9**, there exist a closed set F and an open set G such that $F \subset A \subset G$ and $\lambda(G - F) < \epsilon$. We have $\lambda(G) \leq \lambda(G - F) + \lambda(A) \leq \epsilon + C < \infty$ which yields the contradiction. So we may assume that $\lambda(A) = \infty$. Since $A \cap B(0,k) \in \mathcal{L}_0$, by **M5**,

$$\infty = \lim_{k\to\infty} \lambda(A \cap B(0,k)) = \lim_{k\to\infty} \lambda_*(A \cap B(0,k)) \leq \lambda_*(A)$$

and therefore $\lambda_*(A) = \infty$. $\square$

Proposition 2.20 (**M11**)**.** *If $A \subset B$ and $B \in \mathcal{L}$, then $\lambda^*(A) + \lambda_*(B \setminus A) = \lambda(B)$.*

PROOF. Fix an open set $G \supset A$. Then

$$\begin{aligned}\lambda(G) + \lambda_*(B \setminus A) &\geq \lambda(B \cap G) + \lambda_*(B \setminus A)\\ &\geq \lambda(B \cap G) + \lambda_*(B \setminus G)\\ &= \lambda(B \cap G) + \lambda(B \setminus G) = \lambda(B).\end{aligned}$$

Since G is arbitrary, $\lambda^*(A) + \lambda(B \setminus G) \geq \lambda(B)$.

Now fix a compact set $K \subset B \setminus A$. Then $A \subset B \setminus K$ and

$$\begin{aligned}\lambda^*(A) + \lambda(K) &\leq \lambda^*(B \setminus K) + \lambda(K)\\ &= \lambda(B \setminus K) + \lambda_*(K) = \lambda(B).\end{aligned}$$

Since K can be arbitrarily chosen, $\lambda^*(A) + \lambda_*(B \setminus A) \leq \lambda(B)$. $\square$

Proposition 2.21 (Carathéodory condition, **M12**)**.** *Let $A \in \mathbb{R}^n$. Then $A \in \mathcal{L}$ if and only if*

$$\lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E \cap A^c).$$

PROOF. Suppose that $A \in \mathcal{L}$. Fix an open set $G \supset E$. Then

$$\lambda(G) = \lambda(G \cap A) + \lambda(G \cap A^c) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c).$$

Since G is arbitrary, $\lambda^*(E) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$. But we already have $\lambda^*(E) \leq \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$ from proposition 2.8, ***3**.

Conversely, let $M \in \mathcal{L}_0$. If we choose $E = M$, then from the hypothesis we have

$$\lambda(M) = \lambda^*(M \cap A) + \lambda^*(M \cap A^c).$$

Also, by **M11**, we get

$$\lambda(M) = \lambda^*(M \cap A^c) + \lambda_*(M \setminus (M \cap A^c)) = \lambda^*(M \cap A^c) + \lambda_*(M \cap A).$$

Comparing these two identities and using the fact that $M \in \mathcal{L}_0$, we have

$$\lambda^*(M \cap A) = \lambda_*(M \cap A) < \infty$$

and thus $M \cap A \in \mathcal{L}_0$. Since M is arbitrary, we can conclude that $A \in \mathcal{L}$. $\square$

Remark 2.8. The above proposition gives another definition of measurable set. Several other texts ([**Fol**], [**Roy**]) use the Carathéodory condition as the definition of measurability.

2.3. Miscellany

2.3.1. Symmetries of Lebesgue measure.

The Lebesgue measure in $\mathbb{R}^n$ enjoys a number of symmetries. Firstly, it is translation-invariant. For a measurable set E and $v \in \mathbb{R}^n$, $E+v = \{x+v : x \in E\}$ is also measurable and $\lambda(E+v) = \lambda(E)$. This invariance inherited from the special case when E is a cube and a special polygon. For general sets, since the Lebesgue measure is defined as an approximation of measures of special polygons, it hold true for all measurable set. By the same reason, (and more complicated and tedious proof), one can check the Lebesgue measure is invariant under rotation, reflection, and furthermore, relatively dialation-invariant. In general, we can summarize as follows:

Theorem 2.22. *Let T be an $n \times n$ matrix and $A \subset \mathbb{R}^n$. Then*

$$\lambda^*(TA) = |\det T|\, \lambda^*(A), \quad \textit{and} \quad \lambda_*(TA) = |\det T|\, \lambda_*(A).$$

In particular, if $A \in \mathcal{L}$, then $TA \in \mathcal{L}$ and

$$\lambda(TA) = |\det T|\, \lambda(A).$$

See [**Jon**] Chapter 3 for detail.

2.3.2. Non-measurable set.

We show the existence of non-measurable sets in $\mathbb{R}^n$. The proof is highly nonconstructive, that relies on *the Axiom of Choice.*

Theorem 2.23. *There exists a set $E \subset \mathbb{R}^n$ such that E is not measurable. (i.e. $\mathcal{L} \subsetneq \mathcal{P}(\mathbb{R}^n)$)*

Proof. We will use the translation invariance. For given $x \in \mathbb{R}^n$, consider the translate $x + \mathbb{Q}^n = \{x + r : r \in \mathbb{Q}^n\}$. Crucially, we observe that either

$$x + \mathbb{Q}^n = y + \mathbb{Q}^n \qquad \text{or} \qquad (x + \mathbb{Q}^n) \cap (y + \mathbb{Q}^n) = \varnothing.$$

This means that $\mathbb{R}^n$ is covered disjointly by the translates if $\mathbb{Q}^n$. Now, we invoke the Axiom of Choice to collect exactly one element from each translate of $\mathbb{Q}^n$. Let denote E is the set of collection. Then we have a representation

$$\mathbb{R}^n = \dot{\bigcup_{x\in E}} (x + \mathbb{Q}^n).$$

For another representation, we denote $\mathbb{Q}^n = \{r_1, r_2, \cdots\}$. Then

$$\mathbb{R}^n = \dot{\bigcup_{i=1}^{\infty}} (r_i + E).$$

Since $\lambda(r_i + E) = \lambda(E)$, we conclude that $\lambda^*(E) > 0$. Decomposing $\mathbb{R}^n$ into non-overlapping cubes I_j with sides of length 1, i.e., $\mathbb{R}^n = \bigcup_{j=1}^{\infty} I_j$, we see at least one of $E \cap I_j$'s has positive measure.

We call it $\widetilde{E}$. Then,

$$\bigcup_{r\in\mathbb{Q}^n\cap B(0,1)} r+\widetilde{E}\subset I_j+B(0,1).$$

If $\widetilde{E}$ *were* measurable, due to the countable additivity and the translation invariance of measure, the left-hand side is equal to $\sum_{\text{countable}}\lambda\left(\widetilde{E}\right)$, while $\lambda\left(I_j+B(0,1)\right)<3^n$, which leads contradiction. $\square$

Corollary 2.24. *If $A\in\mathbb{R}^n$ is measurable and $\lambda(A)>0$, then there exists $B\subset A$ such that B is not measurable.*

Proof. Proceed similarly to the previous one with

$$A=\bigcup_{i=1}^{\infty}(r_i+E)\cap A. \qquad \square$$

Using the corollary, one can easily construct a non-measurable set with $\lambda^*(A)=\lambda_*(A)=\infty$.

2.3.3. The Lebesgue function.

Recall the construction of Cantor set. At each step, we remove a third intermediate open interval from each closed intervals. Here, G_k, a removing open set at k-th step, is the union of disjoint open intervals of length 3^{-k} and the number of intervals is 2^{k-1}. Then the leftover C_k is the finite disjoint union of closed intervals of length 3^{-k}. $C_k=[0,1]-\bigcup_{j=1}^k G_j$ and then the Cantor set $C=\bigcap_{k\geq 1}C_k$. Now we will denote $G_k=\bigcup_r J_r$, where J_r is the m-th interval in G_k from the left with $r=\frac{2m+1}{2^k}, m=0,1,\cdots,2^{k-1}-1$. Then,

$$\bigcup_{j=1}^{\infty}G_j=\bigcup_r J_r$$

with $r=\frac{2m+1}{2^k}$ for $0<r<1$. The union is disjoint. We define a function

$$f:\bigcup_{j=1}^{\infty}G_j\to[0,1]$$

so that for each $x\in J_r$, $f(x)=r$. f is constant on each J_r. One can check that f is nondecreasing.

Claim: f is uniformly continuous on its domain.

For a proof, pick $x,y\in G=\bigcup_{j=1}^{\infty}G_j$ with $|x-y|<3^{-k}$. We look at the decomposition of $[0,1]=C_k\cup G_1\cup\cdots\cup G_k$. Since C_k and $G_j, j\leq k$ are the union of disjoint intervals of length $\geq 3^{-k}$, x,y are contained in the same interval or adjacent intervals. Either case,

$$|f(x)-f(y)|\leq f(J_{\frac{m}{2^k}})-f(J_{\frac{m+1}{2^k}})=\frac{1}{2^k}.$$

In general, one can extend a function f to the closure of domain $\overline{G}$. If f is uniformly continuous function, the extended function $\widetilde{f}:\overline{G}\to[0,1]$ is also uniformly continuous. Since $\widetilde{f}$ is constant on G, it is differentiable with $f'=0$ except a measure zero set(=Cantor set). However, the Fundamental Theorem of Calculus fails:

$$1=\widetilde{f}(1)-\widetilde{f}(0)\neq\int_0^1\widetilde{f}'(s)ds=0.$$

In Chapter 4, we will revisit to this example when we investigate the condition on f to hold the Fundamental Theorem of Calculus.

CHAPTER 3

Integration

3.1. algebras and σ-algebra

So far we have constructed $(\mathbb{R}^n, \mathcal{L}, \lambda)$, where λ is the Lebesgue measure

$$\lambda : \mathcal{L} \to [0, \infty].$$

Recall that $\mathcal{L}$ has properties :

(i) $\varnothing \in \mathcal{L}$,
(ii) If $A \in \mathcal{L}$, then $A^c \in \mathcal{L}$,
(iii) If $A_k \in \mathcal{L}$ for $k = 1, 2, \cdots$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{L}$.

And from these properties, we can deduce also that $\bigcap_{k=1}^{\infty} A_k \in \mathcal{L}$ and $\mathbb{R}^n \in \mathcal{L}$.

Definition 3.1 (Algebra and σ-algebra). Denote the power set of X as 2^X or $\mathcal{P}(X)$. Then $\mathcal{M} \subset \mathcal{P}(X)$ is called *algebra* if $\mathcal{M}$ satisfies

(i) $\varnothing \in \mathcal{M}$,
(ii) If $A \in \mathcal{M}$, then $A^c \in \mathcal{M}$,
(iii) If A, $B \in \mathcal{M}$, then $A \cup B \in \mathcal{M}$.

If an algebra $\mathcal{M}$ satisfies the property :

(iii') If $A_k \in \mathcal{L}$ for $k = 1, 2, \cdots$, then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}$,

then we call $\mathcal{M}$ a *σ-algebra*.

Example 3.1. $\mathcal{L}$ is a σ-algebra. The power set $\mathcal{P}(X)$ itself is a σ-algebra. Also, $\{\varnothing, X\}$ forms a σ-algebra.

Proposition 3.1. *Suppose that $\mathcal{M}_i$ is a σ-algebra for all $i \in \mathcal{I}$. Then $\mathcal{M} = \bigcap_{i \in \mathcal{I}} M_i$ is also a σ-algebra.*

Proof. First of all, since $\varnothing \in \mathcal{M}_i$ for all $i \in \mathcal{I}$, $\varnothing \in \mathcal{M}$. Moreover, if $A \in \mathcal{M}$, then $A \in \mathcal{M}_i$ for all $i \in \mathcal{I}$. So $A^c \in \mathcal{M}_i$ for all $i \in \mathcal{I}$ and hence $A^c \in \mathcal{M}$. Finally, if $A_k \in \mathcal{M}$ for $k = 1, 2, \cdots$, then, for each k, $A_k \in \mathcal{M}_i$ for all $i \in \mathcal{I}$. So $\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}_i$ for all $i \in \mathcal{I}$ and therefore we can conclude that $\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}$. □

Now let $\mathcal{N} \in \mathcal{P}(X)$, i.e., $\mathcal{N}$ is a collection of subsets of X. Then we can define

$$\sigma(\mathcal{N}) := \bigcap_{\mathcal{M} \in \Sigma} \mathcal{M},$$

where Σ is the collection of all σ-algebras which contain $\mathcal{N}$. Then from the above proposition, $\sigma(\mathcal{N})$ is a σ-algebra. We say $\sigma(\mathcal{N})$ the *σ-algebra generated by* $\mathcal{N}$. Indeed, $\sigma(\mathcal{N})$ is the smallest σ-algebra containing $\mathcal{N}$.

DEFINITION 3.2 (Borel σ-algebra). Define by $\mathcal{B}_n$(or simply we denote $\mathcal{B}$) the smallest *σ-algebra* containing all open sets in $\mathbb{R}^n$. $\mathcal{B}$ is called the *Borel σ-algebra*. Each element of $\mathcal{B}$ is called a *Borel set*.

Since $\mathcal{L}$ contains all open sets in $\mathbb{R}^n$, we have $\mathcal{B} \subset \mathcal{L}$. Indeed, we will see that $\mathcal{B} \subsetneq \mathcal{L} \subsetneq \mathcal{P}(X)$. In particular, closed sets are Borel sets, and so are all countable unions of closed sets and all countable intersection of open sets. These last two are called F_σ's and G_δ's respectively, and plays a considerable role.[1] With this notation, we can also define $F_{\sigma\delta}$, $G_{\delta\sigma\delta\sigma\delta} \in \mathcal{B}$ and so on.[2]

DEFINITION 3.3. A set $A \in \mathcal{L}$ with $\lambda(A) = 0$ is called a *null set*.

Theorem 3.2. *Let $A \in \mathcal{L}$. Then $A = E \cup N$, where N is a null set, E is a F_σ set, and N and E are disjoint.*

PROOF. See Remark 2.7. For any $k \in \mathbb{N}$, there exists a closed set F_k such that $F_k \subset A$ and $\lambda(A \setminus F_k) < \frac{1}{k}$. Let $E = \bigcup_{k=1}^{\infty} F_k$. Then E is a F_σ set and $\lambda(A \setminus E) = 0$. $\square$

Theorem 3.3. *Let E be a Borel set in $\mathbb{R}^n$. Suppose that a function $f : E \to \mathbb{R}^m$ is continuous. If A is a Borel set in $\mathbb{R}^m$, then $f^{-1}(A)$ is a Borel set.*

PROOF. Define

$$\mathcal{M} := \{A : A \in \mathbb{R}^m \text{ and } f^{-1}(A) \in \mathcal{B}_n\}.$$

We want to show $\mathcal{M}$ is a σ-algebra containing all open sets.

- $f^{-1}(\varnothing) = \varnothing$. Hence $\varnothing \in \mathcal{M}$.
- Suppose $A_k \in \mathcal{M}$, $k = 1, 2, \cdots$. Then $f^{-1}(A_k) \in \mathcal{B}_n$ for $k = 1, 2, \cdots$. Now

$$f^{-1}\left(\bigcup_{k=1}^{\infty} A_k\right) = \bigcup_{k=1}^{\infty} f^{-1}(A_k) \in \mathcal{B}_n.$$

  Therefore, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}$.
- Suppose $A \in \mathcal{M}$. Then $f^{-1}(A) \in \mathcal{B}_n$. Now

$$f^{-1}(A^c) = f^{-1}(\mathbb{R}^m) \setminus f^{-1}(A) = E - f^{-1}(A) \in \mathcal{B}_n.$$

  Thus $A^c \in \mathcal{M}$.

To show that $\mathcal{M}$ contains open sets, we use the continuity of f. By definition, if G is open, then, $f^{-1}(G)$ is open in E, i.e., $f^{-1}(G) = E \cup H$ for some open set H in $\mathbb{R}^n$. So, $f^{-1}(G) \in \mathcal{B}_n$. It implies that $G \in \mathcal{M}$. So all the open sets are contained in $\mathcal{M}$. Finally, we have $\mathcal{B}_m \subset \mathcal{M}$, which completes the proof. $\square$

[1] The notation is due to Hausdorff. The letters F and G were used for closed and open sets (*Fermé* and *Gebeit*), respectively, and σ refers to union (*Summe*), δ to intersection (*Durchschnitt*).

[2] For example, $F_{\sigma\delta}$ is the countable intersection of F_σ's.

Theorem 3.4. $\mathcal{B} \subsetneq \mathcal{L}$

Proof. Let C be a Cantor set on $[0,1]$. Let f be a Lebesgue function on C. We define $g(x) := f(x) + x$ for $0 < x < 1$. Then g is strictly increasing, continuous, and $g(0) = 0$, $g(1) = 1$. Thus $g : [0,1] \to [0,2]$ becomes a homeomorphism. For $x \in J_r$, $g(x) = x + r$. Hence $g(J_r)$ is an open interval of length $\lambda(J_r)$, i.e., $\lambda(g(J_r)) = \lambda(J_r)$. Now we have

$$\lambda(g(C)) = \lambda\left([0,2] \setminus \bigcup_r g(J_r)\right) = 1 > 0.$$

Since $g(C)$ has positive measure, there exists a nonmeasurable set $B \subset g(C)$. Let $A := g^{-1}(B)$. Then $A \subset C$, and hence $\lambda^*(A) \leq \lambda(C) = 0$. Therefore, $A \in \mathcal{L}$. If $A \in \mathcal{B}$, $g(A) = B \in \mathcal{B}$. However, B is not even measurable. Finally, we can conclude that $A \in \mathcal{L}$ but $A \notin \mathcal{B}$. $\square$

Now, we can generalize the Lebesgue measure on $\mathbb{R}^n$ to a general measure on a set X.

Definition 3.4. A measure space is a triple of $(X, \mathcal{M}, \mu)$ as the following:

- X is a nonempty set.
- $\mathcal{M} \subset \mathcal{P}(X)$ is a σ-algebra on X
- $\mu : \mathcal{M} \to [0,\infty]$ is a function satisfying $\mu(\varnothing) = 0$ and

$$\text{if } A_1, A_2, \cdots, \in \mathcal{M} \text{ are disjoint, then } \mu(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k).$$

A measure space $(X, \mathcal{M}, \mu)$(or simply denoted by X or (X, m)) is said to be finite if $\mu(X) < \infty$. If $X = \bigcup_{i=1}^{\infty} X_i$ with $\mu(X_i) < \infty$, then we say X is σ-*finite*.

Remark 3.2. One check that fundamental properties of measures such as the monotonicity, **M4**, **M5**, M6.

Example 3.3.

(1) Lebesgue measure space $(\mathbb{R}^n, \mathcal{L}, \lambda)$.
(2) Borel measure space $(\mathbb{R}^n, \mathcal{B}, \lambda)$.
(3) $(\mathbb{Z}^n, \mathcal{P}(\mathbb{Z}^n), \mathfrak{c})$ where $\mathfrak{c}$ is the counting measure.
(4) Dirac delta measure $(X, \mathcal{M}, \delta_p)$ where p is a point of X. For $A \in \mathcal{M}$,

$$\delta(A) = \begin{cases} 1, & p \in A \\ 0, & p \notin A. \end{cases}$$

We will revisit the general measure theory later in this chapter.

3.2. Measurable Functions

We turn our attention to integrand functions. In order to define a integral, we restrict a natural class of functions on which the integral is well-defined and satisfies fine properties.
We consider the extended real line $[-\infty, \infty]$ and a function defined on X:

$$f : X \to [-\infty, \infty].$$

Let $\mathcal{M}$ be a σ-*algebra* on X. We say f is $\mathcal{M}$-*measurable* if $f^{-1}([-\infty, t]) \in \mathcal{M}$, i.e., $\{x \in X : f(x) \leq t\} \in \mathcal{M}$, for all $t \in [-\infty, \infty]$. If $X = \mathbb{R}^n$, we naturally consider $\mathcal{L}$-measurable functions or $\mathcal{B}$-measurable functions. In short, we say it is a measurable function if f is $\mathcal{L}$-measurable and f is a Borel function if it is $\mathcal{B}$-measurable.

Proposition 3.5. *Let $\mathcal{M}$ be a σ-algebra of a space X. Let f be an extended real-valued function on X. Then the followings are equivalent.*

(i) *f is $\mathcal{M}$-measurable.*
(ii) *$f^{-1}([-\infty, t)) \in \mathcal{M}$ for all $t \in [-\infty, \infty]$.*
(iii) *$f^{-1}([t, \infty]) \in \mathcal{M}$ for all $t \in [-\infty, \infty]$.*
(iv) *$f^{-1}((t, \infty]) \in \mathcal{M}$ for all $t \in [-\infty, \infty]$.*
(v) *$f^{-1}(\{-\infty\})$, $f^{-1}(\{\infty\}) \in \mathcal{M}$ and $f^{-1}(E) \in \mathcal{M}$ for $E \in \mathcal{B}$.*

Proof. Observe that

$$f^{-1}([-\infty, t)) = \bigcup_{\substack{r > t \\ r:\text{rational}}} f^{-1}([-\infty, r]).$$

So (i) implies (ii). And similar observations leads us to the conclusion that statements (i) to (iv) imply each other.

(v) implies (i) trivially. It remains to show that (i) implies (v). First, (i) implies that $f^{-1}(\{-\infty\}) \in \mathcal{M}$. And (iii), which is equivalent to (i), implies $f^{-1}(\{\infty\}) \in \mathcal{M}$. Now, define

$$\mathcal{S} = \{E \in \mathbb{R} : f^{-1}(E) \in \mathcal{M}\}.$$

You can easily check that $\mathcal{S}$ is a σ-algebra. If G is a open set in $\mathbb{R}$, then we can write $G = \dot{\bigcup}_{j=1}^{\infty} I_j$, where $I_j = (a, b) = [-\infty, b) \cup (a, \infty]$. Note that $f^{-1}(I_j) \subset \mathcal{M}$ for each j. So, $I_j \in \mathcal{S}$ for all j and thus $G \in \mathcal{S}$. Hence, $\mathcal{B} \subset \mathcal{S}$ and for any $E \in \mathcal{B}$, $f^{-1}(E) \in \mathcal{M}$. $\square$

Proposition 3.6. *Let f, $g : X \to \mathbb{R}$ be $\mathcal{M}$-measurable functions.*

(**MF 1**) *If $\phi : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $\phi \circ f$ is $\mathcal{M}$-measurable.*

(**MF 2**) *If $f \neq 0$, $\frac{1}{f}$ is $\mathcal{M}$-measurable.*

(**MF 3**) *For $0 < p < \infty$, $|f|^p$ is $\mathcal{M}$-measurable.*

(**MF 4**) *$f + g$ is $\mathcal{M}$-measurable.*

(**MF 5**) *fg is $\mathcal{M}$-measurable.*

(**MF 6**) *Suppose that $f_k : X \to [-\infty, \infty]$ is measurable for all $k \in \mathbb{N}$. Then the following functions are $\mathcal{M}$-measurable.*

$$\sup_k f_k, \quad \inf_k f_k, \quad \limsup_{k \to \infty} f_k, \quad \liminf_{k \to \infty} f_k, \quad \lim_{k \to \infty} f_k \ (\textit{if it exists}).$$

Proof. **MF 4** Fix $t \in \mathbb{R}$. $f(x) + g(x) < t$ if and only if there is a rational number r such that $f(x) < r < t - g(x)$. Therefore,

$$\{x : f(x) + g(x) < t\} = \bigcup_{r \in \mathbb{Q}} f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, t - r)).$$

For **MF 5**, write $fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$ and use **MF 3**.
Others are left as exercise. □

Definition 3.5. Define

$$\chi_A = \mathbf{1}_A := \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

We call χ_A a *characteristic function* (or *indicator function*). Note that $A \in \mathcal{M}$ if and only if χ_A is $\mathcal{M}$-measurable.
A $\mathcal{M}$-measurable simple function $s : X \to [-\infty, \infty]$ is any function which can be expressed in

$$s = \sum_{k=1}^{m} \alpha_k \chi_{A_k}$$

for some $m \in \mathbb{N}$ and $\alpha_k \in \mathbb{R}$, where A_k's are disjoint $\mathcal{M}$-measurable functions.

Remark 3.4. The notion of simple function is more general than *step function*, which is written by

$$s = \sum_{k=1}^{m} c_k \chi_{R_k},$$

where R_k's are nonoverlapping rectangles.

Example 3.5. Let $A = \mathbb{Q} \cap [0,1]$ and $B = [0,1] \setminus A$. Then both $f_{\text{Dir}} = \chi_A$ and χ_B are characteristic function.

For an extended real valued function f from X, there is a way to write f as the difference of two nonnegative functions. First, define

$$f_+(x) = \begin{cases} f(x), & f(x) \geq 0, \\ 0, & f(x) \leq 0 \end{cases} \quad \text{and} \quad f_-(x) = \begin{cases} 0, & -f(x) \geq 0, \\ -f(x), & f(x) \leq 0 \end{cases}$$

Then $f = f_+ - f_-$ and $f_+, f_- \geq 0$.

Theorem 3.7. *Let $f : X \to [-\infty, \infty]$ be a nonnegative [$\mathcal{M}$-measurable] function. Then f can be approximated pointwisely by an increasing sequence of [$\mathcal{M}$-measurable] simple functions.*

Proof. Define

$$s_k(x) = \begin{cases} \dfrac{i-1}{2^k}, & \frac{i-1}{2^k} \leq f(x) < \frac{i}{2^k}, \ i = 1, 2, \cdots, k2^k, \\ k, & f(x) \geq k. \end{cases}$$

Then $\{s_k\}$ is an increasing sequence of ($\mathcal{M}$-measurable) simple functions such that s_k converges to f pointwise. □

We can further approximate by step functions.

Corollary 3.8. *Let $f : X \to [-\infty, \infty]$ be a nonnegative [$\mathcal{M}$-measurable] function. Then f can be approximated almost everywhere by a sequence of step functions.*

PROOF. It suffices to approximate a characteristic function $\mathbf{1}_A$ for a measurable set A. From Theorem 3.2 except for a null set A is G_δ-set, $\bigcap_{k=1}^{\infty} G_k$ where G_k's are open. Each $A_m = \bigcap_{k=1}^{m} G_k$ is a countable union of non overlapping rectangles, $A_m = \bigcup_{j=1}^{\infty} R_m^j$. Then $\{\mathbf{1}_{S_m} : S_m = \bigcup_{j=1}^{m} R_m^j\}$ converges to $\mathbf{1}_A$ a.e. $\square$

In view of $f = f_+ - f_-$, the nonnegative condition is not necessary.

Theorem 3.9. *Suppose that $f : \mathbb{R}^n \to [-\infty, \infty]$ is Lebesgue measurable. Then there exists a Borel measurable function g such that the set $\{x \in \mathbb{R}^n : f(x) \neq g(x)\}$ is a null set.*

PROOF. In view of $f = f_+ - f_-$, we may assume $f \geq 0$. Find a nondecreasing sequence of simple functions $\{s_k\}$ converging to f. For each $s_k = \sum_{j=1}^{m_k} c_j \chi_{A_j^k}$, we replace measurable sets A_j^k by Borel sets B_j^k with $\lambda\left(A_j^k \Delta B_j^k\right) = 0$. Then $\sigma_k = \sum_{j=1}^{m_k} c_j \chi_{B_j^k}$ is a Borel function which agrees its values with s_k except for a null set N_k. Thus, $\{\sigma_k\}$ converges to f pointwise except for $N = \bigcup_{k=1}^{\infty} N_k$, a null set. The limit $\lim_{k\to\infty} \sigma_k = g$ is the Borel function. $\square$

DEFINITION 3.6. If some property is valid except on a null set, we say that the property hold *almost everywhere*, abbreviated a.e.[3] For instance, Theorem 3.9 tells that $f = g$ almost everywhere.

Finally, we introduce two useful theorems.

Theorem 3.10. *(Egorov) Suppose $\{f_k\}_{k=1}^{\infty}$ is a sequence of measurable functions defined on $E \subset \mathbb{R}^n$, $\lambda(E) < \infty$ and assume that $f_k \to f$ a.e. on E. For given $\epsilon > 0$, there exists a closed set $A \subset E$ such that $\lambda(E \setminus A) \leq \epsilon$ and $f_k \to f$ uniformly on A.*

PROOF. Fix $\epsilon > 0$. We may assume $f_k(x) \to f(x)$ for every $x \in E$. Let $n, k \in \mathbb{Z}_+$ and

$$E_k^n = \{x \in E : |f_j(x) - f(x)| < 1/n, \quad \text{for all } j > k\}.$$

For a fixed n, $\{E_k^n\}_{k=1}^{\infty}$ is increasing to E. By M5, we can find k_n so that $\lambda(X \setminus E_k^n) < \frac{1}{2^n}$. Then, we have $|f_j(x) - f(x)| < 1/n$ whenever $j > k_n$ and $x \in E_{k_n}^n$.
We choose N so that $\sum_{n=N}^{\infty} 2^{-n} < \epsilon/2$, and let $B = \bigcap_{n\geq N} E_{k_n}^n$. Then $\lambda\left(E \setminus B\right) \leq \epsilon/2$. On the other hand, $f_j \to f$ uniformly on B. Indeed, for given $\delta > 0$ we choose $n > \max(N, 1/\delta)$. For $x \in B \subset E_{k_n}^n$, $|f_j(x) - f(x)| < \delta$ whenever $j > k_n$. Lastly, using approximation lemma, we can find a closed set $A \subset B$ with $\lambda\left(B \setminus A\right) < \epsilon/2$. $\square$

Theorem 3.11. *(Lusin) Suppose $f : E \to (-\infty, \infty)$ is measurable with $\lambda\left(E\right) < \infty$. Then for given $\epsilon > 0$, there exists a closed set $A \subset E$ satisfying $\lambda\left(E \setminus A\right) < \epsilon$ and $f|_A$ is continuous.*

PROOF. We use Egorov's theorem. Let $\{f_k\}$ be a sequence of step functions converging to f. Then we can find sets E_k so that $\lambda\left(E_k\right) < 2^{-k}$ and f_k is continuous outside E_k. By Egorov's theorem, we can find a set B on which $f_k \to f$ uniformly and $\lambda\left(E \setminus B\right) < \epsilon/3$. Choose N such that $\sum_{k=N}^{\infty} 2^{-k} < \epsilon/3$. Define $A' = B \setminus \bigcup_{k\geq N} E_k$. As f_k is continuous on A' for $k > N$ and $f_k \to f$ uniformly on A', f is continuous on A'. Lastly, we approximate A' by a closed set A. $\square$

[3]Probabilists often say this *almost surely*.

REMARK 3.6. Egorov's theorem and Lusin's theorem hold true in general setting. In general case, A in the conclusions may not be closed. (In fact, a general measure space may not be a topological space.)

3.3. Integration and convergence theorems

To define the Lebesgue integral, we start with a nonnegative, $\mathcal{L}$-measurable, simple function $s = \sum_{k=1}^{m} c_k \chi_{A_k}$ where $0 \le c_k < \infty$. In this case, we define

$$\int s\, d\lambda := \sum_{k=1}^{m} \alpha_k \lambda(A_k),$$

where $\{A_k\}_{k=1,\cdots,m}$ is a finite collection of disjoint $\mathcal{M}$-measurable sets.

We first need to check the well-defineness of the definition. Suppose that s has two different representations. Suppose that $s = \sum_{k=1}^{a} c_k \chi_{A_k} = \sum_{j=1}^{b} d_j \chi_{B_j}$ where $\{A_k\}$ and $\{B_j\}$ are disjoint collections. Decompose sets into $C_{kj} := A_k \cap B_j$, if $C_{kj} = \varnothing$, then $s = c_k = d_j$. We have

$$\int s\, d\lambda = \sum_{k=1}^{a} c_k \lambda(A_k) = \sum_{k,j} c_k \lambda(C_{kj}) = \sum_{k,j} d_j \lambda(C_{kj}) = \sum_{j=1}^{b} d_j m_{B_j}.$$

Proposition 3.12. *Let s, t be simple measurable nonnegative functions.*

- $0 \le \int s\, d\lambda \le \infty$
- *For $c \ge 0$, $\int cs\, d\lambda = c \int s\, d\lambda$*
- $\int (s+t)\, d\lambda = \int s\, d\lambda + \int t\, d\lambda$
- *If $s \le t$, then $\int s\, d\lambda \le \int t\, d\lambda$.*

PROOF. Exercise. □

For a general nonnegative function $f : \mathbb{R}^n \to [0, \infty]$, we define

$$\int f\, d\lambda := \sup\left\{ \int s\, d\lambda : s \le f,\ s : \text{simple, nonnegative, } \mathcal{L}\text{-measurable} \right\}.$$

REMARK 3.7. It is instructive to compare this definition with Riemann integral. In the Riemann integral, the integral is approximated by that of step functions.

For a general measurable function $f : \mathbb{R}^n \to [-\infty, \infty]$, we write $f = f_+ - f_-$ and define its integral by

$$\int f\, d\lambda = \int f_+\, d\lambda - \int f_-\, d\lambda,$$

when both $\int f_+ \lambda$ and $\int f_-\, d\lambda$ are finite. In this case, we say f is integrable (or in L^1).
For a measurable set $E \subset \mathbb{R}^n$, we define the integral on E by

$$\int_E f\, d\lambda := \int f\, \chi_E\, d\lambda.$$

Proposition 3.13. *Let f, g be integrable functions.*

(1) $\int cf\,d\lambda = c\int f\,d\lambda$
(2) $\int (f+g)\,d\lambda = \int f\,d\lambda + \int g\,d\lambda$
(3) *For disjoint measurable sets E, F,* $\int_{E\cup F} f\,d\lambda = \int_E f\,d\lambda + \int_F f\,d\lambda$.
(4) *If $f \leq g$, then* $\int f\,d\lambda \leq \int g\,d\lambda$.
(5) $\left|\int f\,d\lambda\right| \leq \int |f|\,d\lambda$

Proof. We leave (1),(4), and (5) as exercise. The proof of (2) will be given after next theorem. (3) follows from (2). □

We begin to discuss convergence theorems. For monotone sequence of measurable sets $\{A_k : k = 1, 2\cdots\}$ with $A_k \subset A_{k+1}$. we have $\lim_{k\to\infty} \lambda(A_k) = \lambda(A)$ where $A = \bigcup_{k=1}^\infty A_k$. That is,

$$\lim_{k\to\infty}\int \chi_{A_k}\,d\lambda = \int \chi_A\,d\lambda.$$

This is true for a monotone sequence of measurable functions.

Theorem 3.14. *(Monotone Convergence Theorem)*[4] *Assume that $\{f_k : k = 1, 2, \cdots\}$ is increasing sequence of nonnegative measurable functions on $\mathbb{R}^n$. Then*

$$\lim_{k\to\infty}\int f_k\,d\lambda = \int\left(\lim_{k\to\infty} f_k\right)d\lambda.$$

Proof. Denote $f = \lim_{k\to\infty} f_k$ and $I = \lim_{k\to\infty}\int f_k\,d\lambda$. As we have $I \leq \int f\,d\lambda$, we show the other inequality. Fix $c < \int f\,d\lambda$. It suffice to show $I \geq c$. By definition of the integral of f, there exist a simple function s such that $\int s\,d\lambda > c$ and $0 \leq s \leq f$. Let s be of the form $s = \sum_{i=1}^m c_i\chi_{A_i}$ where A_i's are disjoint and measurable. We replace s by a new simple function, still denote by s by changing c_i to $c_i - \epsilon$, where $\epsilon > 0$ is small, so that $c < \int s\,d\lambda$.(Verify!) Then if $f(x) > 0$ then $s(x) < f(x)$. Define $E_k := \{x : f_k(x) \geq s(x)\}$. Then we have $\bigcup_{k=1}^\infty E_k = \mathbb{R}^n$.(Verify!) For a fixed k we have

$$f_k \geq f_k\chi_{E_k} \geq s\chi_{E_k} = \sum_{i=1}^m c_i\chi_{A_i\cap E_k}.$$

Therefore, $\int f_k\,d\lambda \geq \sum_{i=1}^m c_i\lambda(A_i\cap E_k)$. Taking limit of k, $\lim_{k\to\infty}\lambda(A_i\cap E_k) = \lambda(A_i)$, we conclude

$$I = \lim_{k\to\infty}\int f_k\,d\lambda \geq \sum_{i=1}^m c_i\lambda(A_i) = \int s\,d\lambda > c.$$

□

Corollary 3.15. *Let $\{f_k : k = 1, 2, \cdots\}$ be a decreasing sequence of nonnegative measurable functions on $\mathbb{R}^n$. Assuem $\int f_1\,d\lambda < \infty$. Then*

$$\lim_{k\to\infty}\int f_k\,d\lambda = \int\left(\lim_{k\to\infty} f_k\right)d\lambda.$$

PROOF. Use Theorem 3.14 for $\{f_1 - f_k\}_{k=1}^{\infty}$. □

Proof of Proposition 3.13 (2)
First, we prove this for nonnegative functions.
Let $\{s_k\}$ and $\{t_k\}$ be an increasing sequence of simple function converging to f, g, respectively. Then $s_k + t_k$ is increasing to $f + g$. We use additivity property of simple function's integral and MCT to obtain

$$\begin{aligned}\int (f+g)\,d\lambda &= \lim_{k\to\infty} \int (s_k + t_k)\,d\lambda \\ &= \lim_{k\to\infty} \Big[\int s_k\,d\lambda + \int t_k\,d\lambda\Big] \\ &= \int f\,d\lambda + \int g\,d\lambda\end{aligned}$$

For general case, denoting $h = f + g = h_+ - h_- = f_+ - f_- + g_+ - g_-$,

$$h_+ + f_- + g_- = h_- + f_+ + g_+,$$

which implies

$$\int h_+\,d\lambda + \int f_-\,d\lambda + \int g_-\,d\lambda = \int h_-\,d\lambda + \int f_+\,d\lambda \int g_+\,d\lambda.$$

Hence, we conclude $\int h_\pm\,d\lambda < \int f_\pm\,d\lambda + \int g_\pm\,d\lambda < \infty$ and $\int h\,d\lambda = \int f\,d\lambda + \int g\,d\lambda$. □

Corollary 3.16. *(Fatou's Lemma) Assume that $\{f_k : k = 1, 2, \cdots\}$ are nonnegative measurable functions. Then*

$$\int \Big(\liminf_{k\to\infty} f_k\Big)\,d\lambda \le \liminf_{k\to\infty} \int f_k\,d\lambda.$$

PROOF. Define $g_k = \inf\{f_k, f_{k+1}, \cdots\}$. Then $g_k \ge 0$, $g_k \le f_k$, and $\{g_k\}$ is increasing sequence of measurable functions. Using MCT, we obtain

$$\begin{aligned}\int \Big(\liminf_{k\to\infty}\Big) f_k\,d\lambda &= \int (\lim g_k)\,d\lambda \\ &= \lim_{k\to\infty} \int g_k\,d\lambda \\ &\le \liminf_{k\to\infty} \int f_k\,d\lambda.\end{aligned}$$

□

Corollary 3.17. *(Lebesgue's Dominated Convergence Theorem)*[5] *Assume $\{f_k : k = 1, 2, \cdots\}$ is a sequence of measurable functions on $\mathbb{R}^n$ that converge to f a.e. Assume that there exist $g \in L^1$ such that $|f_k(x)| \le g(x)$ a.e.*
Then $f \in L^1$ and

$$\int \Big(\lim_{k\to\infty} f_k\Big)\,d\lambda = \lim_{k\to\infty} \int f_k\,d\lambda.$$

PROOF. We apply Fatou's Lemma to nonnegative functions $g + f_k$, and $g - f_k$. Indeed, we have

$$\int g \pm f \, d\lambda \leq \liminf_k \int (g \pm f_k) \, d\lambda$$

That is,

$$\int \pm f \, d\lambda \leq \liminf_k \int \pm f_k \, d\lambda.$$

Hence,

$$\limsup_k \int f_k \, d\lambda \leq \int f \, d\lambda \leq \liminf_k \int f_k \, d\lambda.$$

□

Corollary 3.18. *Let $\{f_k : k = 1, 2 \cdots\}$ be a sequence of measurable functions on $\mathbb{R}^n$. Assume either $f_k \geq 0$ or $\int \left(\sum_{k=1}^{\infty} |f_k| \right) d\lambda < \infty$. Then*

$$\int \left(\sum_{k=1}^{\infty} f_k \right) d\lambda = \sum_{k=1}^{\infty} \int f_k \, d\lambda.$$

Corollary 3.19. *(Bounded Convergence Theorem) Let $\{f_k : E \to [-M, M]\}$ be a sequence of bounded measurable functions. Assume $\lambda(E) < \infty$ and $f_k \to f$ a.e. Then*

$$\lim_{k\to\infty} \int_E f_k \, d\lambda = \int_E \lim_{k\to\infty} f_k \, d\lambda.$$

REMARK 3.8. MCT, Fatou'lemma, and LCT are almost equivalent. More precisely, one can show Fatou's lemma first and use it to show MCT. If $f \in L^1$ in MCT, it can be proved by LCT.

3.4. Examples

EXAMPLE 3.9.

$$\lim_{n\to\infty} \int_0^n \left(1 - \frac{x}{n}\right)^n x^s \, dx = \int_0^\infty e^{-x} x^s \, dx, \quad (s > -1)$$

Set $f_n(x) = \left(1 - \frac{x}{n}\right)^n \cdot \mathbf{1}_{[0,n]}$. One can observe $\{f_n\}$ is nonnegative increasing sequence converging to e^{-x}. Then use monotone convergence theorem.

EXAMPLE 3.10.

$$\int_0^\infty \frac{\sin x}{e^x - 1} \, dx = \sum_{n=1}^{\infty} \int_0^\infty e^{-nx} \sin x \, dx = \sum_{n=1}^{\infty} \frac{1}{n^2 + 1}$$

Expanding $\frac{1}{e^x - 1}$, use LCT for the first identity.

EXAMPLE 3.11. Consider a double sequence $\{a_{mn}\}_{m,n=1}^{\infty}$. Assume either (i) $a_{mn} \geq 0$ or (ii) $\sum_n \sum_m |a_{mn}| < \infty$. Then

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{mn} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{mn} \tag{3.1}$$

For proof, we understand that the summation over m as an integral over $\mathbb{N}$ with counting measure. Setting $f_n : \mathbb{N} \to \mathbb{R}$ with $f_n(m) = a_{mn}$. Then $\int_{\mathbb{N}} f_n dc = \sum_{m=1}^{\infty} a_{mn}$ and we rewrite (3.1) as

$$\int_{\mathbb{N}} \sum_{n=1}^{\infty} f_n \, dc = \sum_{n=1}^{\infty} \int_{\mathbb{N}} f_n \, dc.$$

One can use either MCT or LCT (or its corollary) to verify the (3.1).
As a corollary of (3.1), we can show *Riemann's rearrangement theorem* as follows. Let $\alpha : \mathbb{N} \to \mathbb{N}$ be a one-to-one correspondence. Assume either (i) $a_n \geq 0$ or (ii) $\sum_{n=1}^{\infty} |a_n| < \infty$. Then

$$\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} a_{\alpha(n)}.$$

For a proof, set $a_{mn} = \delta_n^{\alpha(m)} a_n$.
As an exercise, find an example of $\{a_{mn}\}$ so that (3.1) fails.
Later, we will understand the double summation as a double integral and use Fubini or Toneli theorem to show (3.1) In fact, under an analogous condition, we can switch the order of integral in double integrals.

Example 3.12. Consider a function with two variable $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Integrating over a variable we define

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

Main question here is under what condition we can switch the integral and differentiation. i.e.

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

Since differentiation is defined as a limit, this problem is reduced to see when one can switch the order of limits and integral sign. Thus, we can use convergence theorem obtained in the previous section.
Let $f(x,y)$ be integrable in x-variable. Assume that there exist a dominating function $g(x) \in L^1$ such that $|f(x,y)| \leq g(x)$. Then LCT says that

$$\lim_{y \to y_0} \int f(x,y) \, dx = \int \lim_{y \to y_0} f(x,y) \, dx.$$

In other words, if $f(x,y)$ is continuous with respect to y-variable and satisfies above condition, then $F(y)$ is also continuous.
Next, consider a derivative. Denote $D_2^h f(x,y) := \frac{f(x,y+h) - f(x,y)}{h}$. Then $\lim_{h \to 0} D_2^h f(x,y) = \frac{\partial}{\partial y} f(x,y)$. If $|D_2^h f(x,y)| \leq g(x) \in L^1$, then

$$\frac{d}{dy} F(y) = \lim_h \int D_2^h f(x,y) \, dx = \int \lim_h D_2^h f(x,y) \, dx = \int \frac{\partial}{\partial y} f(x,y) \, dx.$$

In view of mean value theorem, if $|\frac{\partial f}{\partial y}(x,y)| \leq g(x) \in L^1$, then we have the same conclusion.
In fact, the assumption of Lebesgue convergence theorem replaces *uniform convergence* in compact setting. Recall that if $f_n : [a,b] \to \mathbb{R}$ uniformly converges to f, then $\lim_n \int_a^b f_n(x) \, dx = \int f(x) \, dx$.

Example 3.13. Show

$$\int_0^{\infty} \frac{\sin t}{t} \, dt = \frac{\pi}{2}.$$

We understand the integral as an improper integral on $[0, n]$ as $n \to \infty$.
Note that $\lim_{t\to 0} \frac{\sin t}{t} = 1$. One can show this by a contour integral in complex variable (exercise). In this example we use Lebesgue convergence theorem for an alternative proof.
Define $g_n(x) := \int_0^n e^{-tx} \frac{\sin t}{t}\, dt$. Then by direct computation,

$$g_n'(x) = -\int_0^n e^{-tx} \sin t \, dt = -\frac{e^{-nx}(-x\sin n - \cos n) + 1}{1 + x^2},$$

and $|g_n'(x)| \le \frac{e^{-x}(x+1)+1}{1+x^2} \in L^1$. Using LCT,

$$\int \lim_n g_n'(x)\, dx = \lim_n \int g_n'(x)\, dx = \lim_n [g_n(n) - g_n(0)].$$

By computation, we obtain $\int \lim_n g_n'(x)\, dx = -\int_0^\infty \frac{1}{1+x^2}\, dx = -\frac{\pi}{2}$, $\lim_n g_n(0) = \int_0^\infty \frac{\sin t}{t}\, dt$, and $\frac{\sin t}{t} \le 1$, $g_n(n) \le \int_0^n e^{-nt}\, dt \le \frac{1}{n} \to 0$ as $n \to \infty$. This completes the proof.

3.5. A relation to Riemann integrals

In the introduction, we discussed Riemann integrals on $[a, b]$ and its limitation. In many ways we can understand Lebesgue integral is an extension of Riemann integral. Now we discuss Riemann integrability in the context of Lebesgue measure theory. The discussion below works for higher dimension, too. For simplicity, we restrict ourselves to one dimension. Let $f : [a, b] \to \mathbb{R}$ be a bounded function. (Recall that we defined Riemann integral for such a function).

Theorem 3.20. *If f is Riemann integrable, then f is measurable and*

$$\int_{[a,b]}^{\mathcal{R}} f(x)\, dx = \int_{[a,b]}^{\mathcal{L}} f(x)\, d\lambda. \tag{3.2}$$

f is Riemann integrable if and only if f is continuous almost everywhere (= except a null set).

Proof. By definition of Riemann integrability, we can find a sequence of partitions $\{P_k\}$ so that $\lim_k \mathrm{URS}(P_k, f) = \lim_k \mathrm{LRS}(P_k, f) = \int_{[a,b]}^{\mathcal{R}} f(x)\, dx$. For each k, denote $P_k = \{x_0, \cdots, x_N\}$ and step functions

$$s^k(x) = \sum_{i=1}^N \max_{x_{i-1}\le x\le x_i} f(x)\chi_{[x_{i-1},x_i]}, \quad s_k(x) = \sum_{i=1}^N \min_{x_{i-1}\le x\le x_i} f(x)\chi_{[x_{i-1},x_i]}.$$

Then s^k, s_k are measurable function and by definition of Lebesgue integral of simple functions, $\mathrm{URS}(P_k, f) = \int_{[a,b]}^{\mathcal{L}} s^k$ and $\mathrm{LRS}(P_k, f) = \int_{[a,b]}^{\mathcal{L}} s_k$. Furthermore, $s_k(x) \le f(x) \le s^k(x)$. Define $U(x) = \lim_k s^k(x)$ and $L(x) = \lim s_k(x)$. Then using Lebesgue convergence theorem (check the assumption!) and Riemann integrability,

$$\int_{[a,b]}^{\mathcal{R}} f(x)\, dx = \int_{[a,b]}^{\mathcal{L}} U(x)\, d\lambda = \int_{[a,b]}^{\mathcal{L}} L(x)\, d\lambda,$$

and so $U = L$ almost everywhere. Since $L(x) \le f(x) \le U(x)$, $f(x)$ is measurable, $f = U = L$ almost everywhere and

$$\int_{[a,b]}^{\mathcal{L}} f\, d\lambda = \int_{[a,b]}^{\mathcal{R}} f(x)\, dx.$$

For the second statement, we show that f is continuous at x if and only if $L(x) = U(x)$. Then It follows that f is continuous a.e. $\Leftrightarrow L(x) = U(x)$ a.e. $\Leftrightarrow f$ is Riemann integrable, as the second equivalence is verified at the first step.
For 'only if' part, suppose that f is continuous at x. For a fixed $\epsilon > 0$ there exists $\delta > 0$ such that $|f(y) - f(z)| < \epsilon$ for $z, y \in [x - \delta, x + \delta]$. Choose a partition P, the interval containing x of which belongs to $[x - \delta, x + \delta]$. Then $|U(x) - L(x)| \leq |s^P(x) - s_P(x)| \leq \epsilon$. Since ϵ is arbitrary, $U(x) - L(x) = 0$. Conversely, from $U(x) = L(x)$ there exist a partition P such that $|s^P(x) - s_P(x)| \leq \epsilon$. Let $\delta = \min\{|x - z| : z \in P\}$. Then if $|x - y| < \delta$, x, y are in the same interval of partition and hence $|f(x) - f(y) \leq |s^P(x) - s_P(x)| \leq \epsilon$.

$\square$

3.6. Fubini's theorem for $\mathbb{R}^n$

When an integrand has more than one variable, a repeated integral is often an efficient tool to evaluate the higher dimensional integral. In this calculation, we implicitly use the Fubini's theorem. In this section, we discuss Fubini and Tonelli's theorem for a special case, Lebesgue measure on $\mathbb{R}^n$. Without much difficulty this theorem is extended to general measure spaces. We just refer the general case.
Let l, m, and n be dimensions with $l + m = n$. Consider a measurable function $f : \mathbb{R}^m \times \mathbb{R}^l \to [-\infty, \infty]$ with respect to the Lebesgue measure $\mathcal{L}^n$. We denote $f = f(x, y) = f(z)$ where $x \in \mathbb{R}^m$, $y \in \mathbb{R}^l$, and $z = (x, y) \in \mathbb{R}^n$. For a fixed y, $f_y(x) = f(x, y)$ is a function on $\mathbb{R}^m$.

Theorem 3.21. *(Fubini-Tonelli) Let $f : \mathbb{R}^n \to [-\infty, \infty]$ be a measurable function. Assume either f is nonnegative or f is integrable. Then, we have*

- *The function $f_y(x) : \mathbb{R}^m \to [-\infty, \infty]$ is measurable y a.e.*
  The function $f_x(y) : \mathbb{R}^l \to [-\infty, \infty]$ is measurable x a.e.
- *$F(y) = \int f_y(x)\, dx$ is measurable on $\mathbb{R}^l$*
  $G(x) = \int f_x(y)\, dy$ is measurable on $\mathbb{R}^m$
- $$\int_{\mathbb{R}^l} \int_{\mathbb{R}^m} f(x, y)\, dxdy = \int_{\mathbb{R}^n} f(x, y)\, d\lambda(x, y) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^l} f(x, y)\, dydx$$

Proof. By symmetry of x, y variables, we only prove the theorem for $f_y(x), F(y)$. We prove the theorem when f is nonnegative and integrable. If f is integrable, we obtain the same result by writing $f = f_+ - f_-$. If f is just nonnegative, we use MCT to obtain it. We first consider a special case, the characteristic functions. We rephrase it as follows:

Claim: Let $A \subset \mathbb{R}^n$ be a bounded measurable set. Then,

- $A_y := \{x \in \mathbb{R}^m | (x, y) \in A\}$ is measurable a.e. in y.
- $\lambda(A_y)$ is a measurable function of y.
- $\int_{\mathbb{R}^l} \lambda(A_y)\; dy = \lambda(A)$.

Proof of Claim
Step 1 $J \subset \mathbb{R}^n$ is a special rectangle.
Clearly, $J = J_1 \times J_2$ where J_i are special rectangles in $\mathbb{R}^m$ and $\mathbb{R}^l$, respectively. Then,

$$J_y = \begin{cases} J_1, & y \in J_2 \\ \varnothing, & y \notin J_2. \end{cases}$$

Hence, J_y is measurable, $\lambda(J_y) = \lambda(J_1)\,\mathbf{1}_{J_2}(y)$ is a measurable function, and $\int \lambda(J_y)\ dy = \lambda(J_1)\,\lambda(J_2)$.

Step 2 $G \subset \mathbb{R}^n$ is an bounded open set.
G is expressed by a countable union of special rectangles, $G = \dot\bigcup_{k=1}^{\infty} J_k$. Thus, $G_y = \dot\bigcup_{k=1}^{\infty} J_{k,y}$ is measurable as each $J_{k,y}$ is measurable. Moreover, $\lambda(G_y) = \sum_{k=1}^{\infty} \lambda(J_{k,y})$ and

$$\int \lambda(G_y)\ dy = \sum_{k=1}^{\infty} \int \lambda(J_{k,y})\ dy = \sum_{k=1}^{\infty} \lambda(J_k) = \lambda(G),$$

where we used Step 2 in the second equality.

Step 3 $K \subset \mathbb{R}^n$ is a compact set.
We choose a bounded open set $G \supset K$. Then $G \setminus K$ is open. Since $G_y = (G_y \setminus K_y) \,\dot\cup\, K_y$, K_y is measurable and $\lambda(K_y) = \lambda(G_y) - \lambda(G_y \setminus K_y)$ is a measurable function by Step 2. Moreover,

$$\begin{aligned} \int \lambda(K_y)\ dy = \int \lambda(G_y)\ dy - \int \lambda(G_y \setminus K_y)\ dy &= \lambda(G) - \lambda(G \setminus K) \\ = \lambda(G) - \lambda(G) + \lambda(K) &= \lambda(K). \end{aligned}$$

Step 4 $B = \bigcup_{j=1}^{\infty} K_j$ where $\{K_j\}$ is an increasing sequence of compact sets. $C = \bigcap_{j=1}^{\infty} G_j$ where $\{G_j\}$ is an decreasing sequence of bounded open sets.
Since $B_y = \bigcup_{j=1}^{\infty} K_{j,y}$, B_y is measurable, and $\lambda(B_y) = \lim_{j\to\infty} \lambda(K_{j,y})$. By MCT,

$$\begin{aligned} \int \lambda(B_y)\ dy &= \lim_{j\to\infty} \int \lambda(K_{j,y})\ dy \\ &= \lim_{j\to\infty} \lambda(K_j) = \lambda(B) \end{aligned}$$

The case for C is proved similarly to Step 3.

Step 5 General case. $A \subset \mathbb{R}^n$ is a bounded measurable set.
From approximation by open sets and compact sets, we can find a decreasing sequence of open sets, and an increasing sequence of compact sets:

$$K_1 \subset K_2 \subset \cdots \subset A \subset \cdots G_2 \subset G_1,$$

and $\lim_{j\to\infty} \lambda(K_j) = \lambda(A) = \lim_{j\to\infty} \lambda(G_j)$. Define $B = \bigcup_{j=1}^{\infty} K_j$ and $C = \bigcap_{j=1}^{\infty} G_j$. Then $\lambda(B) = \lambda(A) = \lambda(C)$ and $B \subset A \subset C$.

$$\int \lambda(C_y) - \lambda(B_y)\ dy = \lambda(C) - \lambda(B) = 0.$$

Since $\lambda(C_y) - \lambda(B_y) \geq 0$, $\lambda(C_y) - \lambda(B_y) = 0$ a.e. y. Thus, $C_y \setminus B_y$ is a null set, and so is $C_y \setminus A_y$. Hence, A_y is measurable and $\lambda(A_y)$ $(= \lambda(C_y)$ a.e.$)$ is a measurable function a.e. Moreover,

$$\int \lambda(A_y)\, dy = \int \lambda(C_y)\, dy = \lambda(C) = \lambda(A).$$

□

The theorem is valid for simple functions with bounded support as the conclusion is valid for a finite linear combination.

Claim : If the theorem is valid for a increasing sequence of functions $\{f_k \geq 0\}_{k=1}^{\infty}$, then it is valid for its limit function.

Proof of Claim We use MCT. Denote $\lim_{k\to\infty} f_k = f$. As $\{f_{j,y}\}$ is also increasing sequence of measurable functions, f_y is also measurable. Similarly,

$$F(y) = \int f_y(x)\, dx = \lim_{k\to\infty} \int f_{k,y}(x)\, dx = \lim_{k\to\infty} F_k(y),$$

which implies $F(y)$ is a measurable function. Again by MCT and assumption,

$$\int F(y)\, dy = \lim_{k\to\infty} \int F_k(y)\, dy = \lim_{k\to\infty} f_k(x,y)\, d\lambda(x,y) = \int f(x,y)\, d\lambda(x,y).$$

□

When f is nonnegative, from Claim above, considering a sequence of increasing simple functions, we obtain the same conclusion. For general integrable function f, writing $f = f_+ - f_-$ where $f_\pm$ are nonnegative integrable, we have the conclusion. □

This theorem is extended to general measure space without much difficulty. Here we just sketch and state theorem. For more detail we refer Chapter 11 of [**Jon**] or Section 2.5 of [**Fol**].
Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be measure spaces. We define a product measure space on $X \times Y$.

Definition 3.7. A subset of $X \times Y$ of form $A \times B$ where $A \in \mathcal{M}$ and $B \in \mathcal{N}$ are measurable is called a measurable rectangle. We denote $\mathcal{M} \times \mathcal{N} := \sigma(\{A \times B : A \in \mathcal{M},\ B \in \mathcal{N}\})$. That is, $\mathcal{M} \times \mathcal{N}$ is the smallest σ-algebra containing all measurable rectangles.

To construct a product measure space, we need to construct a measure $\pi : \mathcal{M} \times \mathcal{N} \to [0, \infty]$ on a measurable space $(X \times Y, \mathcal{M} \times \mathcal{N})$. Intuitively, for each measurable rectangle, we expect

$$\pi(A \times B) = \mu(A)\, \nu(B).$$

Definition 3.8. We say $(X, \mathcal{M}, \mu)$ is σ- finite if $X = \bigcup_{j=1}^{\infty} A_j$ with $\mu(A_j) < \infty$.

There exists a unique product measure under σ-finite assumption.

Theorem 3.22. *Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be σ-finite measure spaces. Then, for any $E \in \mathcal{M} \times \mathcal{N}$, we have*

(1) *$E_y \in \mathcal{M}$ for all $y \in Y$ and $E_x \in \mathcal{N}$ for all $x \in X$.*

(2) $x \mapsto \nu(E_x)$ *is a* $\mathcal{M}$*-measurable function.*
$y \mapsto \mu(E_y)$ *is a* $\mathcal{N}$*-measurable function.*

(3)
$$\int_Y \mu(E_y)\, d\nu = \int_X \nu(E_x)\, d\mu =: \pi(E)$$

This defines a measure $\pi : \mathcal{M} \times \mathcal{N} \to [0, \infty]$ *satisfying*

$$\pi(A \times B) = \mu(A)\, \nu(B),$$

and such a measure is unique.

Once we have Theorem 3.22, the general Fubini-Tonelli's theorem follows.

Theorem 3.23. *Let* $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ *be* σ*-finite measure spaces and* $(X \times Y, \mathcal{M} \times \mathcal{N}, \pi)$ *is the product measure space. Assume* $f : X \times Y \to [-\infty, \infty]$ *is* $\mathcal{M} \times \mathcal{N}$*-measurable. Furthermore, assume either* f *is nonnegative or integrable. Then we have*

- $f_y(x)$ *is* $\mathcal{M}$*-measurable, and* $f_x(y)$ *is* $\mathcal{N}$*-measurable.*
- *The function* $y \mapsto \int_X f_y(x)\, d\mu$ *is* $\mathcal{N}$*-measurable.*
  The function $x \mapsto \int_Y f_x(y)\, d\nu$ *is* $\mathcal{M}$*-measurable.*
-
$$\int_Y \int_X f_y(x)\, d\mu\, d\nu \int_{X \times Y} f(x, y)\, d\pi = \int_X \int_Y f_x(y)\, d\nu\, d\mu.$$

CHAPTER 4

L^p spaces

4.1. Basics of L^p spaces

Let $(X, \mathcal{M}, \mu)$ be a measure space and let $1 \le p < \infty$. Let $f : X \to [-\infty, \infty]$ be a measurable function. Then $|f|^p$ is also measurable. We define

$$\begin{aligned} 1 \le p < \infty, \qquad & f \in L^p(X, \mu) \quad \Leftrightarrow \quad \int_X |f|^p d\mu < \infty \\ p = \infty, \qquad & f \in L^\infty(X, \mu) \quad \Leftrightarrow \quad \sup\{M : |f(x)| \le M \text{ a.e.}\} < \infty \end{aligned}$$

Note that we regard f as the equivalence class of all functions which are equal to f almost everywhere. Thus, $L^p(\mu)$ is the set of the equivalence class of functions rather than functions. With usual addition and scalar multiplication of functions, we see that $L^p(\mu)$ is a vector space. In fact, $L^p(\mu)$ is a normed space with

$$\begin{aligned} \|f\|_p &= \left(\int_X |f|^p \, d\mu \right)^{1/p} \\ \|f\|_\infty &= \sup\{M : |f(x)| \le M \text{a.e.}\} = \inf\{t : \mu(|f(x)| > t)\} \end{aligned}$$

We need to check the conditions for a norm $\| \cdot \|_p$

(1) If $f \in L^p$, then $\|f\|_p < \infty$.
(2) $f = 0$ a.e. $\Leftrightarrow \|f\|_p = 0$.(This is the main reason that we understand an element of L^p is an equivalence class.)
(3) $\|cf\|_p = |c|\|f\|_p$
(4) (Triangle inequality) $\quad \|f + g\|_p \le \|f\|_p + \|g\|_p$

Readers can easily check (1) $\sim$ (3). Triangle inequality which is often referred as Minkowski's inequality will be proven later. For simplicity of notation, denote simply L^p for Lebesgue measure. When the measure space is $(\mathbb{Z}^n, \mathcal{P}, \mathfrak{c})$, we denote $L^p(\mathfrak{c}) = l^p$.

Remark 4.1. (Complex valued functions)
We can extend our integration theory to complex valued functions without any difficulty. For given $f : X \to \mathbb{C}$, writing $f(x) = \operatorname{Re} f(x) + \operatorname{Im} f(x)$, we say f is measurable if and only if $\operatorname{Re} f$ and $\operatorname{Im} f$ are measurable. Define L^p-norm as usual

$$\|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p}.$$

Then, it is easy to check $f \in L^p$ if and only if $\operatorname{Re} f, \operatorname{Im} f \in L^p$.
In view of Fourier transform, it is natural to work on complex valued functions. Fortunately, the integration theory extends without further difficulty. One thing you should aware of is that in complex plane, functions cannot take ∞ as a value. In other words, in the extended real line,

$\sup_n |f_n|$ exists always in $[-\infty, \infty]$ but $\sup_n f_n(x)$ may not exist in complex plane. This may cause a little inconvenience, but when one work with L^p functions, there is no problem since $|f(x)| < \infty$ almost everywhere.

REMARK 4.2. For any measurable function f, one can find a Borel function g such that $f = g$ a.e. Hence, for L^p-functions, we do not see that difference between measurable functions and Borel functions. We can simply assume they are Borel functions.

EXAMPLE 4.3. Consider the Lebesgue measure space $(\mathbb{R}^n, \mathcal{L}, \lambda)$ with $p < q$. There are several notion of convergence with respect to each norm.

(1) A sequence which converges in L^p but does not converge in a.e.
Write $n = 2^j + r$ with $0 \le r < 2^j$. Define $f_n = \chi_{[r2^{-j},(r+1)2^{-j}]}$. Then $\|f_n\|_p = 2^{-j/p}$ but the sequence does not converge to zero at every points in $[0, 1]$.

(2) A sequence which converges in a.e. but does not converge in a.e.

$$f_n = n^{1/p}\chi_{[0,\frac{1}{n}]}, \qquad \|f_n\|_p = 1$$

(3) A sequence which converges in L^p but does not converges in L^q.

$$f_n = n^{1/q}\chi_{[0,\frac{1}{n}]}, \qquad \|f_n\|_q = 1, \quad \|f_n\|_p = n^{\frac{p}{q}-1}$$

This counter example employs a set with arbitrarily small positive measure.

(4) A sequence which converges in L^q but does not converges in L^p.

$$f_n = n^{-1/p}\chi_{[0,n]}, \qquad \|f_n\|_p = 1, \quad \|f_n\|_q = n^{-\frac{q}{p}+1}$$

This counter example employs a set with arbitrarily large measure.

EXAMPLE 4.4. In the measure space which does not have arbitrarily large set or arbitrarily small set in measure, we have an inclusion between L^p.

(1) $L^p(\mathbb{Z}, c) = l^p$
This measure space with counting measure does not have sets of arbitrarily small positive measure. For $p < q$, we have $l^p \subset l^q$. Let $\{a_n\} \in l^p$. Then

$$\begin{aligned}\sum_n |a_n|^q &= \sum_{|a_n|\le 1} |a_n|^q + \sum_{|a_n|>1} |a_n|^q \\ &\le \sum_{|a_n|\le 1} |a_n|^p + \sum_{|a_n|>1} |a_n|^q \\ &\le \|\{a_n\}\|_p + C < \infty\end{aligned}$$

where we used $\{n : |a_n| > 1\}$ is finite because of $\{a_n\} \in l^q$.

(2) $L^p(X, \mu)$ with $\mu(X) < \infty$, on which there is no arbitrarily large set. For $p < q$, we have $L^q \subset L^p$.
Let $f \in L^q$. Then

$$\begin{aligned}\int_X |f|^p \, d\mu &= \int_{|f|\le 1} |f|^p \, d\mu + \int_{|f|>1} |f|^p \, d\mu \\ &\le \mu(X) \cdot 1 + \int_{|f|>1} |f|^q \, d\mu < \infty.\end{aligned}$$

4.2. Completeness of L^p spaces

One of the most important gain of the Lebesgue theory is the completeness of L^p function spaces.(You may recall that a limit of Riemann integrable functions may not even Riemann integrable.)

Theorem 4.1. *(Rietz-Fischer) Let $1 \leq p \leq \infty$. L^p is a complete normed space.*

In general, a complete normed vector space is called a *Banach space*. For a proof, we need to verify triangle inequality and the completeness. To proceed for the proof, we begin with the *Hölder's* inequality. This can be viewed as a generalization of the *Cauchy-Schwartz* inequality.

Proposition 4.2. *(Hölder's inequality)*
Let $1 \leq p, q, r \leq \infty$ with $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$. Suppose $f \in L^p$ and $g \in L^q$. Then $fg \in L^r$ and

$$\|fg\|_r \leq \|f\|_p \|g\|_q$$

Proof. Hölder inequality is a consequence of the convexity of exponentials, which formulated as Young's inequality: Let $a, b > 0$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then,

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q.$$

To verify this, set $A = a^p, \quad B = b^q$ and $t = \frac{1}{p}, \quad 1 - t = \frac{1}{q}$. Then we reduce to show

$$(\frac{A}{B})^t \leq t\frac{A}{B} + (1-t),$$

which is followed by the convexity of $f(t) = \alpha^t$.
Back to the Hölder inequality, we may assume that $r = 1$(considering $|f|^\theta, |g|^\theta$) and $f, g \geq 0$ but f, g are nonzero functions. (Check!) If $p = \infty$ or $q = \infty$, then the inequality is easy to verify. Assume $1 < p, q < \infty$, $r = 1$.
By Young's inequality, we have

$$\int f(x)g(x)\,d\mu \leq \frac{1}{p}\int |f(x)|^p\,d\mu + \frac{1}{q}\int |g(x)|^q\,d\mu.$$

For any $\lambda > 0$ we can apply the above inequality to $\lambda f(x)$ and $g(x)/\lambda$ to obtain

$$\int f(x)g(x)\,d\mu \leq \frac{\lambda^p}{p}\int |f(x)|^p\,d\mu + \frac{1}{q\lambda^q}\int |g(x)|^q\,d\mu.$$

Optimizing the right hand side by λ, we reach to the Hölder's inequality.[1]
One can observe that the equality holds if and only if $\alpha|f(x)|^p = \beta|g(x)|^q$ a.e. for some α, β.
Alternatively, before applying Young's inequality, one can normalize $\|f\|_p = \|g\|_q = 1$ (by replacing by f, g by $f/\|f\|_p$, $g/\|g\|_q$ if needed). Then, Young's inequality gives

$$\int f(x)g(x)\,d\mu \leq \frac{1}{p}\int |f(x)|^p\,d\mu + \frac{1}{q}\int |g(x)|^q\,d\mu. = \frac{1}{p} + \frac{1}{q} = 1.$$

□

[1] This trick is sometimes called *arbitrage*. One can make use of the symmetry of inequality to reduce an inequality into a weaker inequality.

REMARK 4.5. The scaling condition $\frac{1}{r} = \frac{1}{p} + \frac{1}{q}$ is a necessary condition when the measure μ has a scaling invariance. (eg. Lebesgue measure on $\mathbb{R}^d$) As an exercise, readers can check $\frac{1}{r} \geq \frac{1}{p} + \frac{1}{q}$ for l^p, but $\frac{1}{r} \leq \frac{1}{p} + \frac{1}{q}$ for $L^p([0,1])$.

In fact, L^2 has a richer structure, an inner product,

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

from which $\|\cdot\|_2$ is generated. Using Hölder's inequality, one shows $\langle f, g \rangle$ is finite for $f, g \in L^2$. So, L^2 is a complete inner product vector space that is usually called a *Hilbert space.*

Proposition 4.3. *(Minkowski's inequality)*
For $1 \leq p \leq \infty$, we have

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p$$

PROOF. For $p = \infty$, it is easy to deduce from $|f(x) + g(x) \leq |f(x) + |g(x)|$. Assume $1 \leq p < \infty$. One can observe that $|f(x) + g(x)|^p \leq (|f(x)| + |g(x)|)|f(x) + g(x)|^{p-1}$.

$$\begin{aligned}\int |f+g|^p \, d\mu &\leq \int (|f| + |g|)|f+g|^{p-1} \, d\mu \\ &\leq \|f\|_p \||f+g|^{p-1}\|_q + \|g\|_p \||f+g|^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p)\Big(\int |f+g|^{(p-1)q}\Big)^{1/q}\end{aligned}$$

Using $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\|f+g\|_p \leq \|f\|_p + \|g\|_q.$$

□

Proof of Rietz-Fischer's theorem
We are left to show the completeness. For given a Cauchy sequence $\{f_n\}$ in L^p, we can extract a subsequence such that

$$\|f_{n_{k+1}} - f_{n_k}\|_p \leq 2^{-k}.$$

We denote $f_{n_k} =: f_k$ for simplicity of notation.

Lemma 4.4.

$$\|\sum_{k=1}^{\infty} f_k\|_p \leq \sum_{k=1}^{\infty} \|f_k\|_p$$

PROOF. Exercise. Use Minkowski's inequality and MCT. □

Define $F(x) = |f_1(x)| + \sum_{k=1}^{\infty} |f_{k+1}(x) - f_k(x)|$. Then by the lemma, $F \in L^p$ and so $F(x) < \infty$ a.e. (i.e. $F(x) < \infty$ for $x \in N^c$ with $\mu(N) = 0$.)
Define the limit function

$$f(x) = \begin{cases} f_1(x) + \sum_{k=1}^{\infty} f_{k+1}(x) - f_k(x), & x \in N^c \\ 0, & x \in N \end{cases}.$$

Then, we have

$$f(x) - f_k(x) = \sum_{j=k}^{\infty} f_{j+1}(x) - f_j(x), \qquad \text{a.e.}$$

$$\|f - f_k\|_p \le \sum_{j=k}^{\infty} \|f_{j+1} - f_j\|_p \le 2^{-k+1}$$

Hence, $f_k \to f$ in L^p. □

As a byproduct of the proof, we obtain a useful corollary.

Corollary 4.5. *If a sequences $f_n \to f$ in L^p, the there exists a subsequence $\{f_{n_k}\}$ converging to f almost everywhere.*

4.3. Approximation of L^p functions

Next, we discuss approximation of L^p-functions by smooth functions in L^p-norm. This is an analogue of Stone-Weierstrauss theorem in a compact setting, i.e., a continuous function is approximated by polynomial in the uniform norm.

Theorem 4.6 (Density theorem). *Let $1 \le p < \infty$. $C_c^\infty(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R} : f \in C^\infty$ and $\operatorname{supp} f$ is compact.$\}$ is dense in $L^p(\mathbb{R}^d)$, i.e., for any $\epsilon > 0$ and $f \in L^p$, there exists $g \in C_c^\infty(\mathbb{R}^d)$ such that $\|f - g\|_p < \epsilon$.*

Proof.

1. We take two steps. First, we approximate f by a continuous function with compact support and then approximate the continuous function by a smooth function.

For a fixed $\epsilon > 0$, let $\epsilon_1 = \epsilon/10$. We use MCT to choose $R > 0$ so that $\|f\|_{L^p(B_R^c)} \le \epsilon_1$ since $\|f\|_{L^p(\mathbb{R}^d)} < \infty$. Moreover, we choose $M > 0$ such that $\|f \mathbf{1}_{\{|f|>M\}}\|_p \le \epsilon_1$ also by MCT. We can approximate f by $f\mathbf{1}_{B_R \cap \{|f| \le M\}}$ in L^p. From this observation, we may assume supp $f \in B_R$ and $|f| \le M$.

Now, we approximate f by a simple function s such that $|f(x) - s(x)| \le \frac{\epsilon_1}{\lambda(B_R)}$ a.e in B_R and s is supported in B_R. Then, $\|f - s\|_{L^p(B_R)} \le \epsilon_1$. Say $s = \sum_{k=1}^N \alpha_k \mathbf{1}_{A_k}$. By approximation theorem of measurable sets, we find open sets G_k and closed sets F_k such that $F_k \subset A_k \subset G_k$ with $\lambda(G_k \setminus F_k) \le \frac{\epsilon_1^p}{M}$ where $M = N \max\{\alpha_k\}_{k=1}^N$. Now, we can use Uryson's lemma, to find a continuous function $h_k(x)$ so that $0 \le h_k(x) \le 1$, $\operatorname{supp} h_k \subset G_k$ and $h_k(x) = 1$ on F_k. Set $\widetilde{s} = \sum_{k=1}^{\mathbb{N}} \alpha_k h_k$. Then $\widetilde{s}$ is a continuous function supported on B_{R+1} satisfying

$$\begin{aligned}\|s - \widetilde{s}\|_p &\le \sum_{k=1}^N |\alpha_k| \|\mathbf{1}_{A_k} - h_k\|_p \\ &\le \sum_{k=1}^N |\alpha_k| \lambda(G_k \setminus F_k)^{1/p} \le \epsilon_1\end{aligned}$$

Hence, combining all together, we obtain $\|f - \widetilde{s}\|_p < \epsilon/2$.

2. For simplicity of notation, assume $f \in L^p$ is continuous and supported in B_R.

Before getting into the proof we prepare two things, which are independently useful for many applications.

Convolution

The convolution of two sequences naturally appears in the product of two power series expansion. Let $\{a_n\}_{n=0}^\infty$ and $\{b_n\}_{n=0}^\infty$ be sequences. We define the convolution $a*b$ by

$$(a*b)_n = a_0b_n + a_1b_{n-1} + \cdots + a_nb_0 = \sum_{j=0}^n a_jb_{n-j}.$$

Then when $f(x) = \sum_{n=0}^\infty a_nx^n$ and $g(x) = \sum_{n=0}^\infty b_nx^n$, the power series expansion of fg is $f(x)g(x) = \sum_{n=0}^\infty (a*b)_nx^n$. In particular, if f, g are periodic functions with period 2π, then one has Fourier series expansion, $f(x) = \sum_{n\in\mathbb{Z}} \widehat{f}(n)e^{inx}$, and $g(x) = \sum_{n\in\mathbb{Z}} \widehat{g}(n)e^{inx}$. Then formally, we obtain

$$\widehat{fg}(n) = (\widehat{f}*\widehat{g})(n), \qquad \widehat{f}(n)\widehat{g}(n) = \widehat{f*g}(n),$$

where $(f*g)(x) = \int_0^{2\pi} f(y)g(x-y)\,dy$.

We extend the convolution to functions on $\mathbb{R}^n$. For f, g measurable functions, we define a convolution by

$$(f*g)(x) := \int f(x-y)g(y)\,dy$$

whenever the integral is well-defined. (i.e $\int |f(x-y)g(y)|\,dy < \infty$ a.e. in x) (eg. $f \in L^p$ and $g \in L^q$) One can easily check the commutativity and associativity for good functions (i.e. $f, g, h \in C_c^\infty$).

$$f*g = g*f \qquad \text{and} \qquad (f*g)*h = f*(g*h)^2.$$

The convolution is useful to modify a rough function to a regular function. If $f \in L^1$, $g \in C^1$ with $|\partial_{x_i} g| \le M$, then using Lebesgue convergence theorem we have $f*g \in C^1$ and

$$\partial_{x_i}(f*g)(x) = (f*\partial_{x_i}g)(x)$$

Approximation of identity

Consider a smooth function $\phi(x)$ satisfying

- $\phi(x) \ge 0$,
- $\phi \in C_c^\infty$ with supp $\phi \subset B_1$,
- $\int \phi(x)\,dx = 1$.

Define a rescaled function $\phi_t(x) = \frac{1}{t^d}\phi(\frac{x}{t})$. Then supp $\phi_t \in B_t$ and $\int \phi_t = 1$. As $t \to 0$, the support of ϕ_t gets smaller but its integral is preserved. However, $\lim_{t\to 0}\phi_t$ does not exist as a measurable function.[3] We call ϕ_t an *approximation of identity* in view that for a measurable function f, we have

$$\lim_{t\to 0} \phi_t * f(x) = f(x)$$

[2]Proof for associativity requires Fubini theorem

[3]The limit is in fact the Dirac delta measure. So, there is no limit as a function. One you show ϕ_t converge to δ_0 in a weaker sense of limit

in several senses. For a full account of property of this, see, for instance, [**Fol**] pp.242.

To finish the proof, we want to show $\|f - \phi_t * f\|_p \leq \epsilon/2$ for small t. Since f is a compactly supported continuous function, f is uniformly continuous, i.e. there is $\delta > 0$ so that

$$|f(x) - f(x-y)| \leq \epsilon_2 \qquad \text{whenever} \quad |y| \leq \delta$$

$$\begin{aligned}
|f(x) - \phi_t * f(x)| &= |f(x) - \int \phi_t(y) f(x-y)\, dy| \\
&\leq \int_{|y|<t} \phi_t(y) |f(x) - f(x-y)|\, dy \\
&\leq \epsilon_2 \qquad \text{choosing} \quad t < \delta \\
\|f - \phi_t * f\|_p^p &\leq \int_{B_R} \left|f(x) - \phi_t * f(x)\right|^p dx \\
&\leq \lambda\left(B_R\right) \epsilon_2^p.
\end{aligned}$$

Choosing $\epsilon_2 = \epsilon/2\lambda\left(B_R\right)$, show $\phi_t * f =: g \in C_c^\infty$ is an approximation of f in L^p norm. □

Exercise 4.1. *Combining Theorem 4.6 and Stone-Weierstrauss theorem, show that L^p is separable for $1 \leq p < \infty$. To the contrary, show that $L^\infty([0,1])$ is not separable.*

4.4. Duality of L^p

In the linear algebra course, we have learned a dual space V^* of a vector space V of finite dimension. V^* is a vector space of linear functionals, i.e. $V^* = \{L : V \to \mathbb{R} |\, L \text{ is linear}\}$. Due to a basis of V, we can characterise V^* and verify $V \equiv V^*$. In infinite dimensional spaces, not every linear functional is bounded and so cannot give a norm of V^*. A natural analogue of linear functionals are *bounded linear functionals.*
Let B be a Banach space. A linear functional $L : B \to \mathbb{C}$ is bounded if $\sup_{x\neq 0} \frac{|L(x)|}{\|x\|} =: \|L\| < \infty$. Then, $B^* = \{L : B \to \mathbb{C} \,|\, L \text{ is a bounded linear functional } \}$ is a normed space with $\|L\|$.(Check!)
Exercise 4.2. *Show B^* is a Banach space for any normed space B. (One has to use the completeness of $\mathbb{C}$.)*
Exercise 4.3. *Let $L : B \to \mathbb{C}$ be a linear functional for a normed space B. The followings are equivalent:*

(1) *L is continuos.*
(2) *L is continuous at* 0.
(3) *L is bounded.*

Now we discuss the duality of $L^p(\mu)$ spaces for $1 \leq p \leq \infty$.
Let (p, q) be a conjugate pair, i.e. $\frac{1}{p} + \frac{1}{q} = 1$. For each $g \in L^q$, one define a linear functional on L^p by

$$L_g(f) = \int fg d\mu.$$

Due to the Hölder's inequality, L_g is bounded linear functional with $\|L_g\| \leq \|g\|_q$. In fact, when $1 \leq q < \infty$, $\|L_g\| = \|g\|_q$ by choosing $f \in L^p$ such that $fg = |g|^q$ a.e. To be more precise, choose $f = 0$ where $g = 0$ and if $g \neq 0$, then $f = |g|^{q-1}\overline{\text{sgn } g}$. Then we obtain $\int fg\, d\mu = \|g\|_q^q$ and $\|f\|_p = \|g\|_q^{q-1}$, which gives equality of Hölder's inequality. This provides an isometry $L^q \to (L^p)^*$.

When $q=\infty$, we assume μ is the Lebesgue measure.[4] For given $\epsilon>0$, let $A=\{x:|g(x)|\geq \|g\|_\infty-\epsilon\}$. Then $\lambda(A)>0$. Pick a measurable set $B\subset A$ with $\lambda(B)<\infty$. Define $f=\chi_B\overline{\text{sgn }g}$. Then $\|f\|_1=\lambda(B)$ and

$$\|L_g\|\geq\int fg\,d\lambda\|f\|_1^{-1}=\lambda(B)^{-1}\int_B|g|\geq\|g\|_\infty-\epsilon.$$

Since $\epsilon>0$ is arbitrary, $\|L_g\|=\|g\|_\infty$. Therefore, we have L^q is isometrically embedded into $(L^p)^*$.

Theorem 4.7. *Let (p,q) be a conjugate pair with $1\leq p<\infty$. Assume μ is σ-finite. Then, L^q is isometrically isomorphic to $(L^p)^*$.*

Proof. From the above discussion, we are left to show that for a given $L\in(L^p)^*$, there exists a $g\in L^q$ such that $L=L_g$.

Step 1

We may reduce to the case μ is finite. Indeed, write $X=\cup X_i$ where $\mu(X_i)<\infty$. Assume that we have Theorem for X_i. Then there is $g_i\in L^q(X_i)$ such that for $f_i\in L^p(X_i)$, $L(f_i)=\int f_ig_i\,d\mu$. Define $g=\sum_{i=1}^\infty g_i$. Since supp g_i are disjoint, we have $L(f)=\sum_{i=1}^\infty\int f\mathbf{1}_{X_i}g_i\,d\mu=\int fg\,d\mu$. In order to show that $g\in L^q$, we use the boundedness of L. When $1<p<\infty$, choose $f=|g|^{q-1}\text{sgn }g$. Then $\|f\|_p^p=\|g\|_q^q=L(f)$. Setting $Y_n=\bigcup_{i=1}^n X_i$, we have $\frac{|L(f\mathbf{1}_{Y_n})|}{\|f\mathbf{1}_{Y_n}\|_p}=\|f\mathbf{1}_{Y_n}\|_p^{p-1}\leq\|L\|$. Thus, $\|f\|_p<\infty$ and so $g\in L^q$.

When $p=1,q=\infty$, Consider a set $E=\{x\in X:|g(x)|>\|L\|+1\}$ and a subset F with $0<\mu(F)<\infty$. Chooing $f=\mathbf{1}_F\text{sgn }g$, we estimate $L(f)\geq\int|g|\,d\mu\geq(\|L\|+1)\mu(F)=(\|L\|+1)\|f\|_1$, which makes a contradiction.

Step 2

We may reduce to the case L is positive. (i.e. $L(f)\geq 0$ for any nonnegative $f\in L^p$). Indeed, we have

Claim For any $L\in(L^p)^*$, we have a unique decomposition $L=L_+-L_-$, where $L_\pm$ are positive definite.

Proof. We say a measurable set E is totally positive if $L(\mathbf{1}_F)\geq 0$ for any $F\subset E$. Set $M:=\sup_E L(\mathbf{1}_E)\geq 0$ where the supremum is taken over all totally positive sets. Then there exists a sequence of totally positive sets $\{E_k\subset X:k=1,2,\cdots\}$ such that $L(\mathbf{1}_{E_k})\to M$. Then $X_+=\bigcup_{k=1}^\infty E_k$ is a maxima, i.e. $L(\mathbf{1}_{X_+})=M$. (Check!) Then it is easy to check L_+ is positive definite.(first, do it for nonnegative simple functions) Letting $L_-=L_+-L=-L(\cdot\,\mathbf{1}_{X\setminus X_+})$, we have to show that L_- is positive definite. Suppose not. Then, there exist a set E_1 in X_- such that $L(\mathbf{1}_E)>0$. If E_1 is totally positive, then replacing X_+ by $X_+\cup E_1$ makes a contradiction. Thus, E_1 must contain a subset F_1 such that $L(\mathbf{1}_{E_1\setminus F_1})>L(\mathbf{1}_{E_1})$. We choose F_1 such that $L(\mathbf{1}_{E_1\setminus F_1})>L(\mathbf{1}_{E_1})+1/n_1$, and $E_2:=E_1\setminus F_1$, where n_1 is the smallest integer for which such E_2 exists. E_2 cannot be totally positive due to the same argument. Then we repeat to pick $E_3\subset E_2$ such that $L(\mathbf{1}_{E_3})>L\mathbf{1}_{E_2}+1/n_2$, where n_2 is the smallest for which such E_3 exists. Continuing this procedure we construct a decreasing sequence $E_1\supset E_2\supset\cdots\supset X_-$ such that

[4]In general, we need the "semi-finiteness" assumption on measures. See [**Fol**] for general discussion.

$L(\mathbf{1}_{E_{k+1}} > L(\mathbf{1}_{E_k} + 1/n_k$. Then $E := \bigcap_{k=1}^{\infty}$ is totally positive since E cannot contain any subset F such that $L(\mathbf{1}_F) > L(\mathbf{1}_E$. As $L(\mathbf{1}_E) > 0$, it contradicts to the choice of X_+. □

Now, we assume that L is positive definite and μ is finite. We consider the collection of measurable functions

$$\mathcal{S} = \{0 \le g \in L^q : \int f\mathbf{1}_E \, d\mu \le L(\mathbf{1}_E) \quad \text{for all } E \in \mathcal{M}\}, \quad K = \sup_{g\in\mathcal{S}} \int g \, d\mu$$

Then, the maximum is attained for some measurable function $\overline{g} \in \mathcal{S}$ by MCT and the fact that if $g_1, g_2 \in \mathcal{S}$, then $\max\{g_1, g_2\} \in \mathcal{S}$. (Check!)
For a fixed $\epsilon > 0$, we define $L_\epsilon(f) = L(f) - \inf \overline{g} f \, d\mu - \epsilon \int f \, d\mu$. We claim that L_ϵ is negative definite for all $\epsilon > 0$. Assume not. Then, by Claim (decomposition) for some ϵ, $L_{\epsilon +} \neq 0$, and so there exist a non measure zero set E such that $L_{\epsilon +}(\mathbf{1}_F) \ge 0$. That is, $L(\mathbf{1}_F) \ge \int \overline{g}\mathbf{1}_F \, d\mu + \epsilon \int \mathbf{1}_F \, d\mu$ for any $F \subset E$. Then we can replace $\overline{g}$ by $g' = \overline{g}\mathbf{1}_{X\setminus E} + (\overline{g} + \epsilon)\mathbf{1}_E$ and obtain $g' \in \mathcal{S}$ but $\int g' \, d\mu > M$, which makes a contradiction. Next, we can verify that $L(f) = \int f\overline{g} \, d\mu$ for $f \in L^p$. (first, show this for nonnegative simple functions and use the continuity of L and MCT to show it for nonnegative functions, and then write $g = g_+ - g_-$.)
Finally, we show that $\overline{g} \in L^q$. The argument is similar to Step 1, except that we do cut off function value instead of its support. Indeed, for $1 < p < \infty$ define $g_N = \min(\overline{g}, N)$ and $f_N = |g_N|^{q-1}\mathrm{sgn}\, g$. Then, $L(f_N) \ge \int |g_N|^q \, d\mu = \|f_N\|_p^{p-1}\|f_N\|_p$. Since $\|L\|$ is bounded, $\|f_N\|_p^{p-1}$ is bounded in N, and so is $\|g_N\|_q$. Hence, by MCT, $g \in L^q$. When $p = 1, q = \infty$, one can also argue as Step 1.(Exercise!). □

When $p = \infty$, it is known that $(L^\infty)^* \supsetneq L^1$. For more discussion, see [**Fol**] p.191.
Exercise 4.4. *Show the uniqueness of the decomposition in Claim and uniqueness of $\overline{f} \in L^q$.*

Corollary 4.8. *For $1 < p < \infty$, $L^p(\mu)$ is a reflexive Banach space. i.e. $(L^p)^{**} = L^p$.*

The proof of Theorem 4.7 is similar to the discussion of signed measure and the proof of the Radon-Nikodym theorem.

Definition 4.1. Let $(X, \mathcal{M})$ be a measurable space. A signed measure is a map $\mu : \mathcal{M} \to [-\infty, \infty]$ such that

- $\mu(\varnothing) = 0$
- μ can take either the value ∞ or $-\infty$ but not both,
- If $E_1, E_2, \cdots \subset X$ are disjoint, then $\sum_{k=1}^{\infty} \mu(E_k)$ converges to $\mu(\bigcup_{k=1}^{\infty} E_k)$, with the former sum being absolute convergent if the latter expression is finite.

Exercise 4.5. *Let $(X, \mathcal{M}, \mu)$ be a signed measure space. Then, there exists a decomposition of $X = X_+ \cup X_-$ such that $\mu|_{X_+} \ge 0$ and $\mu|_{X_-} \le 0$. That is, for any $E \subset X_+[X_-]$, $\mu(E) \ge [\le]0$. Argue that this decomposition is unique up to null sets. This decomposition deduce a decomposition of signed measure into unsigned measures. Indeed, there is unique decomposition $\mu = \mu_+ - \mu_-$ where $\mu_\pm$ are unsigned measure.*

Exercise 4.6 (Radon-Nikodym)**.** *Let μ be an absolute continuous signed measure to the Lebesgue measure λ. (i.e. If $\lambda(E)=0$, then $\mu(E)=0$) Then, there exists $f \in L^1(\mathbb{R}^n,\lambda)$ such that*

$$\mu(E)=\int_E f\,d\lambda.$$

4.5. More useful inequalities

We discuss some more useful inequalities.

Proposition 4.9. *(log-convexity inequality) Let $f:\mathbb{R}^n \to \mathbb{R}$. and $1 \le p \le r \le q \le \infty$ and $\frac{1}{r}=\theta\frac{1}{p}+(1-\theta)\frac{1}{q}$. Then,*

$$\|f\|_r \le \|f\|_p^{\theta}\|f\|_q^{1-\theta}$$

Proof.
1. Use Hölder's inequality for $f=|f|^{\theta}$ and $g=|f|^{1-\theta}$.
2. Directly show the convexity of $s \to \log\|f\|_s$ by taking the second derivatives.
3. (Tensor power trick) By scaling we may assume that $\|f\|_p=\|f\|_q=1$. First, it is easy to verify

$$\|f\|_r \le 2\|f\|_p^{\theta}\|f\|_q^{1-\theta} \tag{4.1}$$

by dividing into two cases $|f|\ge 1$ or $|f|<1$. Observe that the coefficient 2 is independent of dimension n. We use this symmetry of the inequality to the dimension. Set $\underbrace{f\otimes\cdots\otimes f}_{k}=f^{\otimes k}:\mathbb{R}^{nk}\to\mathbb{R}$ by $f^{\otimes k}(x_1,\cdots,x_k)=f(x_1)f(x_2)\cdots f(x_k)$ where $x_j\in\mathbb{R}^n$. Then, we have (4.1) for $f^{\otimes n}$:

$$\|f^{\otimes k}\|_{L^r(\mathbb{R}^{nk})} \le 2\|f^{\otimes k}\|_{L^p(\mathbb{R}^{nk})}^{\theta}\|f^{\otimes k}\|_{L^q(\mathbb{R}^{nk})}^{1-\theta}.$$

A computation shows that $\|f^{\otimes k}\|_{L^p(\mathbb{R}^{nk})}=\|f\|_{L^p(\mathbb{R}^n)}^k$, hence, we obtain

$$\|f\|_{L^r(\mathbb{R}^n)} \le \sqrt[k]{2}\|f\|_{L^p(\mathbb{R}^n)}^{\theta}\|f\|_{L^q(\mathbb{R}^n)}^{1-\theta}.$$

Since k is arbitrary, we conclude the result. □

Exercise 4.7. *Prove the Hölder's inequality by the tensor power trick.*

Theorem 4.10. *(Minkowski's inequality(integral form))*
Let $f(x,y)$ be a measurable function on $\mathbb{R}^{m+l}$. Then, for $1\le p\le\infty$,

$$\Big(\int\Big(\int f(x,y)\,dx\Big)^p dy\Big)^{1/p} \le \int\Big(\int |f(x,y)|^p\,dy\Big)^{1/p} dx. \tag{4.2}$$

Proof. If $p=1$ it is merely Fubini's theorem and if $p=\infty$, then it is a simple consequence of integral. Assume $1<p<\infty$. We use the dual formulation. Let q is the conjugate exponent and $g\in L^q(y)$.

$$\begin{aligned}\int\Big|\int f(x,y)\,dx\Big||g(y)|dy &\le \int\int |f(x,y)||g(y)|\,dy\,dx\\ &\le \int \|g\|_q\|f(x,\cdot)\|_p\,dx\end{aligned}$$

Using

$$\sup_{\|g\|_q=1} \int fg = \|f\|_p,$$

we conclude (4.2). □

Theorem 4.11. *(Young's inequality)*
For $1 \le p \le \infty$, we have

$$\|f * g\|_p \le \|f\|_p \|g\|_1. \tag{4.3}$$

More generally, suppose that $1 \le p, q, r \le \infty$ with $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$.
If $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, then $f * g \in L^r(\mathbb{R}^n)$ and

$$\|f * g\|_r \le \|f\|_p \|g\|_q. \tag{4.4}$$

Proof. To prove (4.3), we use (4.2)

$$\begin{aligned} \Big(\int \Big| \int f(x-y) g(y)\, dy \Big|^p dx \Big)^{1/p} &\le \int \Big(\int |f(x-y) g(y)|^p \, dx \Big)^{1/p} dy \\ &\le \|f\|_p \|g\|_1. \end{aligned}$$

For the general case (4.4), we may assume $f, g \ge 0$ and normalize f, g so that $\|f\|_p = \|g\|_q = 1$. Using the Hölder's inequality, and observing exponents numerology.

$$\begin{aligned} f * g(x) &= \int f(x-y)^{p/r} g(y)^{q/r} f(x-y)^{1-p/r} g(y)^{1-q/r}\, dy \\ &\le \Big(\int f(x-y)^p g(y)^q \, dy \Big)^{1/r} \Big(\int f(x-y)^{(1-p/r)q'} \Big)^{1/p'} \\ &\times \Big(\int g(y)^{(1-q/r)p'} \Big)^{1/q'} \\ &= \Big(\int f(x-y)^p g(y)^q \, dy \Big)^{1/r} \cdot 1 \cdot 1. \end{aligned}$$

Hence, $\quad (f * g(x))^r \le f^p * g^q(x)$

Using (4.3) for $p = 1$,

$$\int (f * g)^r \, dx \le \int f^p * g^q \, dx \le \|f^p\|_1 \|g^q\|_1 = \|f\|_p^p \|g\|_q^q = 1.$$

Note that we have used the translation invariance of measure in the proof. Generally, Young's inequality holds true for translation invariant measure, so called a *Haar measure*. □

CHAPTER 5

Differentiation

The differentiation and integration are inverse operations. Indeed, the fact is formulated as the Fundamental Theorem of Calculus. More precisely, if $f : [a, b] \to \mathbb{R}$ is a continuous function, then its primitive function:

$$F(x) := \int_a^x f(y)\, dy,$$

is continuously differentiable with $F' = f$. On the other hands, if F is differentiable, then

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

The goal of this chapter is to extend this relation of integral and differentiation to more general functions, Lebesgue measurable functions. We will also extend the analogous statement to higher dimension.
We formulate two questions.

Question 1. Let f be an integrable function and define $F(x) = \int_a^x f(y)dy$. Is f differentiable? If so, under what condition on f do we guarantee $F' = f$?

It is easy to see that F is continuous (actually a bit stronger than continuous). We already know the answer is yes if f is continuous or piecewise continuous. The questions turns to a limiting question of averaging operator:

$$\frac{1}{2h}(F(x+h) - F(x-h)) = \frac{1}{2h} \int_{x-h}^{x+h} f(y)\, dy \to f(x) \quad \text{as } h \to 0?$$

We will study this question in general dimension as follows:

$$\frac{1}{\lambda\,(B_r(x))} \int_{B_r(x)} f(y)\, d\lambda y \to f(x) \quad \text{as } r \to 0?$$

Next question is to find a mild sufficient condition for the Fundamental Theorem of Calculus.

Question 2. What condition on F guarantee that F' exist a.e., that F' is integrable, and that

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

In Chapter 2, we have seen an example, the Lebesgue-Cantor function (See Subsection 2.3.3), for which F' exists and is equal to zero a.e. but Question 2 fails. Hence, we need a stronger condition, so called *absolute continuity*, than just continuity.

5.1. Differentiation of the Integral

5.1.1. Hardy-Littlewood maximal function.

We begin with a geometric lemma.

Theorem 5.1 (Vitali's Covering Lemma). *Let $E \subset \mathbb{R}^n$ be a bounded set. Let $\mathcal{F}$ denote a collection of open balls which are centered at points of E such that every point of E is the center of some ball in $\mathcal{F}$, i.e., $\mathcal{F} = \{B_{r(x)}(x) : x \in E\}$. Then there exists a countable (at most) subcollection $\{B_1, B_2, \cdots\}$ in $\mathcal{F}$ such that B_j's are disjoint and $E \subset \bigcup_{j=1}^{\infty} 3B_j$.*

Proof. Without loss of generality, we may assume radii of balls are bounded. Inductively, assume that $B_1, \cdots, B_{n-1}$ are selected. Let

$$d_n = \sup\{\operatorname{rad} B : B \in \mathcal{F} \text{ and } B \cap \bigcup_{j=1}^{n-1} B_j = \varnothing\}.$$

If there are no $B \in \mathcal{F}$ such that $\bigcup_{j=1}^{n-1} B_j = \varnothing$, then we stop the procedure. Otherwise, we choose $B_n \in \mathcal{F}$ such that B_n is disjoint with $\bigcup_{j=1}^{n-1} B_j$ and $\frac{1}{2} d_n \leq \operatorname{rad} B_n$. The selection gives a countable subcollection $\{B_1, B_2, \cdots\}$ in $\mathcal{F}$ such that B_j's are disjoint.

Pick x in E and let $B = B_{r(x)}(x)$.

Claim : B has a nonempty intersetion with at least one of the balls B_1, B_2, $\cdots$.

If not, this process never terminates. In deed, $r(x) < d_n$ for any $n = 1, 2, \cdots$. However, $d_n \to 0$, since E is a bounded set (of finite measure). Let $\alpha \geq 1$ be the smallest number such that B_α intersects with B. Hence

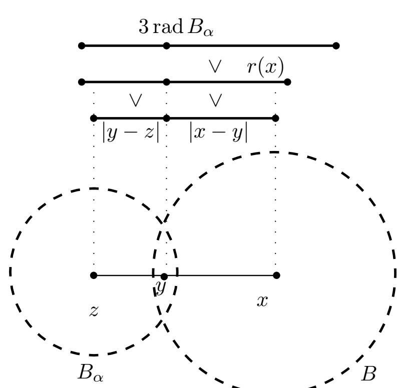
$$B \cap \bigcup_{j=1}^{\alpha-1} B_j = \varnothing$$

and we can conclude that

$$r(x) \leq d_\alpha < 2\operatorname{rad} B_\alpha.$$

Let $y \in B \cap B_\alpha$. Then if z is the center of B_α, we have

$$\begin{aligned} |x - z| &\leq |x - y| + |y - z| \\ &< r(x) + \operatorname{rad} B_\alpha < 3\operatorname{rad} B_\alpha. \end{aligned}$$



Therefore, x must lie in $3B_\alpha$. $\square$

Corollary 5.2. *Let $E \subset \mathbb{R}^n$ be a bounded set. Let $\mathcal{F}$ denote a collection of open balls which are centered at points of E such that every point of E is the center of some ball in $\mathcal{F}$, i.e., $\mathcal{F} = \{B_{r(x)}(x) : x \in E\}$. Then, for any $\epsilon > 0$, there exists $\{B_1, \cdots, B_N\}$ in $\mathcal{F}$ such that B_j's are disjoint and*

$$\lambda\left(E - \bigcup_{j=1}^{N} B_j\right) < \epsilon.$$

DEFINITION 5.1. Let f be a function in $L^1_{\text{loc}}(\mathbb{R}^n)$, i.e., $\int_K |f| < \infty$ for any compact set $K \in \mathbb{R}^n$. Then we define the *Hardy-Littlewood maximal function*

$$Mf(x) = \sup_{0<r<\infty} \frac{1}{B(x,r)} \int_{B(x,r)} |f(y)| dy.$$

The followings are basic properties of the Hardy-Littlewood maximal function:

- Mf is measurable.
- M is sublinear, that is, $M(f+g) \leq Mf + Mg$.
- If f is continuous, $Mf(x) \geq f(x)$. Later, we will see that $Mf(x) \geq f(x)$ a.e. for $f \in L^1_{\text{loc}}(\mathbb{R}^n)$.
- In general, $Mf \notin L^1(\mathbb{R}^n)$ even though $f \in L^1(\mathbb{R}^n)$. In fact, if $Mf \in L^1(\mathbb{R}^n)$, then $f = 0$. Let $a > 0$. For $|x| > a$, we have

$$\begin{aligned} Mf(x) &\geq \frac{1}{\lambda(x, 2|x|)} \int_{B(x,2|x|)} |f(y)|\, dy \\ &\geq \frac{1}{\lambda(0, 2|x|)} \int_{B(0,a)} |f(y)|\, dy \\ &= \frac{1}{V_n\, |x|^n} \int_{B(0,a)} |f(y)|\, dy. \quad (V_d : \text{the volume of the } n\text{-dimensional unit ball}) \end{aligned}$$

  Fix a. Then

$$\int_{|x|>a} Mf(x)\, dx = \int_{B(0,a)} |f(y)|\, dy \int_{|x|>a} \frac{1}{|x|^d}\, dx = \infty,$$

  unless $\int_{B(0,a)} |f(y)|\, dy = 0$. Therefore, if $Mf \in L^1(\mathbb{R}^n)$, then $f \equiv 0$ a.e. on $B(0,a)$. Since a can be chosen arbitrarily, $f \equiv 0$ a.e. on $\mathbb{R}^n$.

EXAMPLE 5.1 (A function $f \in L^1$ such that $Mf \notin L^1_{\text{loc}}$). Let

$$f(x) = \begin{cases} \frac{1}{x \log^2 x} & 0 < x < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Then $f \in L^1(\mathbb{R})$, since

$$\int_0^{\frac{1}{2}} \frac{1}{x \log^2 x} dx \overset{\log x = t}{=} \int_{-\infty}^{\log \frac{1}{2}} \frac{1}{t^2} dt < \infty.$$

However,

$$Mf(x) \geq \frac{1}{2x} \int_0^{2x} f(y) dy > \frac{1}{2x} \int_0^{x} \frac{1}{y \log^2 y} dy = -\frac{1}{2x \log x} \notin L^2_{\text{loc}}(\mathbb{R}).$$

Theorem 5.3. *If $f \in L^1(\mathbb{R}^n)$, then*

$$t\,\lambda(\{x : Mf(x) > t\}) \le 3^n \|f\|_{L^1}$$

for $0 < t < \infty$.

PROOF. Let $E = \{x : Mf(x) > t\}$. For $k \in \mathbb{N}$, denote $E_k = E \cap B(0,k)$, which is bounded. Then for each $x \in E$, there exists an open ball B_x around x such that $t < \frac{1}{\lambda(B_x)} \int_{B_x} |f(y)| dy$. Denote $\mathcal{F} = \{B_x : x \in E_k\}$. Using Vitali's covering lemma, we can extract disjoint sets $B_1, B_2, \cdots$ from $\mathcal{F}$ so that $E_k \subset \bigcup_{j=1}^{\infty} 3B_j$. Then

$$\begin{aligned}
\lambda(E_k) &\le \sum_{i=1}^{\infty} \lambda(3B_j) = \sum 3^n (B_j) \\
&< \sum_{j=1}^{\infty} 3^n t^{-1} \int_{B_j} |f(y)|\, dy = 3^n t^{-1} \int_{\bigcup_{j=1}^{\infty} B_j} |f(x)|\, dx \\
&\le 3^n t^{-1} \int_{\mathbb{R}^n} |f(x)| dx.
\end{aligned}$$

Letting $k \to \infty$, we obtain $t\,\lambda(\{x : Mf(x) > t\}) \le 3^n \|f\|_{L^1}$. □

REMARK 5.2. Theorem 1.1 is a weaker version of the statement (4) in the following sense: We denote

$$\|f\|_{L^{1,\infty}} = \sup_{0<t<\infty} t\,\lambda(\{x : |f(x)| > t\}). \quad (\le \|f\|_{L^1})$$

Then $\|\cdot\|_{L^{1,\infty}}$ defines a norm and we call this norm the *weak* L^1-norm. So M is a bounded operator from L^1 to weak L^1. Moreover, it is easy to check that $M : L^\infty \to L^\infty$.

REMARK 5.3. (optional) As we have the boundedness of M from $L^1 \to L^{1,\infty}$, and $L^\infty \to L^\infty$, one can use real interpolation theorem to conclude

$$\|Mf\|_{L^p} \le C_p \|f\|_{L^p}$$

for some constant C_p depending on p. Therefore, we can conclude that M is a bounded sublinear operator from L^p to L^p, $1 < p \le \infty$. See [**Fol**], [**Tao**] for detail.

5.1.2. Lebesgue's Differentiation Theorem. Now we are ready to answer to *Question 1.*

Theorem 5.4 (Lebesgue's differentiation theorem)**.** *Suppose $f \in L^1_{loc}(\mathbb{R}^n)$. Then*

(1) *for a.e. $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} |f(y) - f(x)|\, dy = 0.$$

(2) *In particular, for a.e. $x \in \mathbb{R}^n$,*

$$\lim_{r\to 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} f(y)\, dy = f(x).$$

PROOF. We first show (2). By replacing f by $f \cdot \mathbf{1}_{R+1}$ and considering $|x| < R$, we may assume that $f \in L^1(\mathbb{R}^n)$. (2) is true if f is a continuous function. Let g be a continuous function such that $\|f - g\|_{L^1} < \epsilon$. Denote $A_r f = \lim_{r\to 0} \frac{1}{\lambda(B(x,r))} \int_{B(x,r)} f(y)\, dy$. Then

$$\begin{aligned}\limsup_{r\to 0} |A_r f(x) - f(x)| &= \limsup_{r\to 0} |A_r(f-g)(x) + (A_r g - g)(x) + (g-f)(x)| \\ &\le M(f-g)(x) + |f-g|(x)\end{aligned}$$

Let

$$E_t = \{x : \limsup_{r\to 0} |A_r f(x) - f(x)| > t\} \quad \text{and} \quad F_t = \{x : |f-g|(x) > t\}.$$

Then $E_t \subset F_{t/2} \cup \{x : M(f-g)(x) > t/2\}$ and

$$\begin{aligned}\lambda(E_t) &\le \lambda(F_{t/2}) + \lambda(\{x : M(f-g)(x) > t/2\}) \\ &\le \frac{\epsilon}{t/2} + \frac{3^n \epsilon}{t/2}.\end{aligned}$$

Since ϵ is arbitrary, $\lambda(E_t) = 0$ for all $t > 0$. Therefore, $\lim_{r\to 0} A_r f(x) = f(x)$ for every $x \notin \bigcup_{n=1}^{\infty} E_{1/n}$, i.e., $x \in \mathbb{R}^n$ a.e.

Now we continue to prove (1). For $x \in \mathbb{C}$, let $g_c(x) = |f(x) - c|$. From (2),

$$\lim_{r\to 0} \lambda(B(x,r)) \int_{B(x,r)} |f(y) - c|\, dy = |f(x) - c|$$

for $x \notin G_c$ with $\lambda(G_c) = 0$. Let D be a countable dense subset of $\mathbb{C}$. (for example, complex numbers with rational coordinates) Then $E = \bigcup_{c\in D} G_c$ is a null set. If $x \notin E$, there exists $c \in \mathbb{C}$ such that $|f(x) - c| < \epsilon$ and so $|f(x) - f(y)| < |f(y) - c| + \epsilon$. Hence

$$\begin{aligned}\lim_{r\to 0} \lambda(B(x,r)) \int_{B(x,r)} |f(x) - f(y)|\, dy &< \lim_{r\to 0} \lambda(B(x,r)) \int_{B(x,r)} |f(y) - c|\, dy + \epsilon \\ &= |f(x) - c| + \epsilon \le 2\epsilon.\end{aligned}$$

Since ϵ is arbitrary,

$$\lim_{r\to 0} \lambda(B(x,r)) \int_{B(x,r)} |f(x) - f(y)|\, dy = 0. \qquad \square$$

Corollary 5.5. *Let $E \subset \mathbb{R}^n$ be a measurable set. Then*

$$\lim_{r\to 0} \frac{\lambda(E \cap B(x,r))}{\lambda(B(x,r))} = \mathbf{1}_E(x) \text{ a.e.} \tag{5.1}$$

REMARK 5.4. The left hand side of (5.1) is often referred as the density of E at x.

DEFINITION 5.2. Suppose $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then the set

$$\{x \in \mathbb{R}^n : \text{there exists } A \text{ such that } \lim_{r\to 0} \lambda(B(x,r)) \int_{B(x,r)} |f(y) - A|\, dy = 0\}$$

is called the *Lebesgue set* of f.

REMARK 5.5. In the Lebesgue's differentiation theorem, using a ball is not crucial. We can generalize that result to the case when a sequence $\{E_k\}$ of measurable sets 'shrinking nicely'. More precisely, we say that $\{U_n\}_{n=1}^{\infty}$ shrink nicely to x if $x \in U_n$, $\lambda(U_n) \to 0$, and there exists a constant $c > 0$ such that for each n there exist a ball B with

$$x \in B, \qquad U_n \subset B, \qquad \text{and} \quad U_n \geq c\lambda(B).$$

Thus, U_n is contained in B but its measure is comparable.

We can generalize the averaging operator to weighted averaging operator.

Exercise 5.1. *Let $f \in L^p(\mathbb{R}^n)$. Let ϕ_t be an approximation of identity. That is, ϕ is radial, nonnegative, $\phi(x) = \phi(|x|) \geq c > 0$ for $|x| \leq 1$ and supp $\phi \in B_2$. Set $\phi_t(x) = t^{-n}\phi(x/t)$. Then, we have*

$$\lim_{t\to 0} f * \phi_t(x) = f(x), \qquad a.e.$$

5.2. Differentiability of Functions

Now, we turn to the second question, that is , finding a sufficient condition on F to guarantee

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

5.2.1. Functions of bounded variation.

DEFINITION 5.3. Let $F(x)$ be a real valued function defined on $[a,b]$. For a partition $x_0 = a < x_1 < \cdots < x_N = b$, we say $\sum_{k=1}^{N} |F(x_k) - F(x_{k-1})|$ is a variation. We say F is of bounded variation ($f \in BV$) if the variations for any partition is bounded by a constant M. Furthermore, we denote the supreme over all partitions is the total variation.

$$T_F(x) = \sup \sum_{k=1}^{N} |F(x_k) - F(x_{k-1})|.$$

EXAMPLE 5.6.

(1) If $f : [a,b] \to \mathbb{R}$ is monotone, then $f \in BV$ and $T_f(x) = f(x) - f(a)$.
(2) If $f \in C^1$, then $f \in BV$. However, there is a continuous function $f : [a,b] \to \mathbb{R}$ which is not BV.

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & 0 < x \leq 1 \\ 0, & x = 0. \end{cases}$$

(3) Let $F(X) = \int_z^x f(y)\, dy$ where f is integrable. Then F is continuous and of bounded variation. Furthermore,

$$T_F(x) = \int_a^x |f(y)|\, dy.$$

Indeed, for any partition $\{x_0, x_1, \cdots, x_N\}$,

$$\sum_{k=1}^{N} |F(x_k) - F(x_{k-1})| = \sum_{k=1}^{N} \int_{x_{k-1}}^{x_k} f(y)\, dx \leq \sum_{k=1}^{N} \int_{x_{k-1}}^{x_k} |f(y)|\, dx \leq \int_a^b |f(y)|\, dy.$$

Hence, we have $T_F(x) \le \int_a^x |f(y)|\,dy$ and $F \in BV$. For the other side of inequality, we approximate f by a step function s such that $\|f-s\|_1 \le \epsilon$. One can check the equality holds true for step functions. Then,

$$T_F(x) \ge T_S(x) - T_{F-S}(x) \le \int_a^x |s(y)|\,dy - \epsilon \ge \int_a^x |f(y)|\,dy - 2\epsilon.$$

(4) Any BV function is bounded and Riemann integrable. But, as seen above, not every Riemann integrable function is BV.

Theorem 5.6. *A real-valued function F on $[a,b]$ is of bounded variation if and only if F is the difference of two increasing bounded functions.*

PROOF. 'if' part is obvious. We prove the 'only if' part. Assume that $F \in BV$. Define $g(x) = T_F(x) - f(x)$. It suffices to show that $g(x)$ is increasing. For $x < y$,

$$\begin{aligned} g(y) - g(x) &= T_F(y) - F(y) - T_F(x) + F(x) \\ &= [T_F(y) - T_F(x)] - [F(y) - F(x)] \\ &\ge F(y) - F(x) - [F(y) - F(x)] = 0 \end{aligned}$$

□

This theorem tells that it suffices to consider monotone functions when studying bounded variation functions.
Now, we show a deep theorem due to Lebesgue.

Theorem 5.7 (Lebesgue)**.** *Functions of bounded variation are differentiable a.e.*

For a proof, we need to use a form of Vitali's covering lemma.

Lemma 5.8 (Vitali's covering lemma, infinitesimal version)**.** *Let $E \subset \mathbb{R}^n$. Let $\mathcal{F}$ be a collection of closed balls with positive radii such that for $x \in E$ and $\epsilon > 0$, there exists a ball $B \in \mathcal{F}$ containing x with $\mathrm{rad}\, B < \epsilon$. Then there exists a countable subcollection $\{B_1, B_2, \cdots\}$ such that B_j's are disjoint and $E \subset \bigcup_{\alpha=1}^{\infty} B_\alpha$ except for a null set.*

PROOF. We may assume that E is bounded. We may further assume that all the balls in $\mathcal{F}$ have less than some positive constant. Moreover, we discard balls in $\mathcal{F}$ which are disjoint with E. Suppose $B_1, \cdots, B_\alpha$ have been selected. If $E \subset \bigcup_{\beta<\alpha} B_\beta$, then we stop our procedure. Otherwise, let

$$d_\alpha = \sup\{\mathrm{rad} B : B \cap \cup_{k=1}^{\alpha} = \varnothing\}$$

□

PROOF (Proof of Theorem). It suffices to show theorem for increasing function f on a bounded interval $[a,b]$. Let

$$Df(x) = \limsup_{\delta \to 0} \left\{ \frac{|f(y) - f(z)|}{y - z} : a \le y \le x \le z \le b,\ 0 < z - y < \delta \right\},$$

$$df(x) = \liminf_{\delta \to 0} \left\{ \frac{|f(y) - f(z)|}{y - z} : a \le y \le x \le z \le b,\ 0 < z - y < \delta \right\}.$$

Then $0 \le df(x) \le Df(x) \le \infty$. We want to show that

$$\{x : Df(x) = \infty\} \cup \{x : df(x) < Df(x)\} = \bigcap_{k \ge 1} \{x : Df(x) > k\} \cup \bigcup_{\substack{s<t \\ s,t \in \mathbb{Q}}} \{x : df(x) < s < t < Df(x)\}$$

is a null set. Let $E = \{x : Df(x) > k\}$ and $F = \{x : df(x) < s < t < Df(x)\}$.

Claim 1. $\lambda^*(E) = \frac{c}{k}$ for some c.

If $x \in E$, then there exists arbitrarily small interval $x \in [y, z] \subset [a, b]$ containing x such that $\frac{f(z)-f(y)}{z-y} > k$. Let $I = [y, z]$ and $\widetilde{I} = (f(y), f(z))$. Then $\lambda(\widetilde{I}) > k\lambda(I)$. We collect all such closed intervals. Then it satisfies the condition for Vitali's covering lemma, so we can find an at most countable subcollection of disjoint intervals $\{I_\alpha\}_{\alpha=1}^{\infty}$ such that $E \subset \bigcup_{\alpha \ge 1} I_\alpha$ a.e. Since f is increasing, $\{\widetilde{I}_\alpha\}$ are also disjoint. We estimate

$$\lambda(E) \le \lambda\left(\bigcup_{\alpha \ge 1} I_\alpha\right) = \sum_{\alpha=1}^{\infty} \lambda(I_\alpha) \le \frac{1}{k} \sum_{\alpha=1}^{\infty} \lambda\left(\widetilde{I}_\alpha\right) \le \frac{f(b) - f(a)}{k}$$

Claim 2. $\lambda^*(F) = 0$.

We will show $\lambda(F) \le \frac{s}{t}\lambda(F)$ and so conclude $\lambda(F) = 0$. For given ϵ we choose an open set $G \supset F$ such that $\lambda(G \setminus F) \le \epsilon$. For the first inequality, for each $x \in F$ we can find arbitrarily small intervals $x \in [y, z]$ such that $[y, z] \subset G \cap [a, b]$ and $\frac{f(y)-f(z)}{y-z} < s$. By Vitali's covering lemma, we can find a collection of disjoint intervals $\{I_\alpha\}$ such that $F \subset \bigcup_{\alpha \ge 1} I_\alpha$ a.e. Thus, we estimate

$$\lambda\left(\bigcup_{\alpha \ge 1} \widetilde{I}_\alpha\right) = \sum_{\alpha \ge 1} \lambda\left(\widetilde{I}_\alpha\right) < s \sum_{\alpha \ge 1} \lambda(I_\alpha) = \lambda\left(\bigcup_{\alpha \ge 1} I_\alpha\right) \le s\lambda(G) \le s(\lambda(F) + \epsilon)$$

For the other side of inequality, we begin with $F \cap \bigcup_{\alpha \ge 1} I_\alpha^\circ$. Note that since I_α's are disjoint closed intervals , $\lambda\left(\bigcup_{\alpha \ge 1} I_\alpha^\circ\right) = \lambda\left(\bigcup_{\alpha \ge 1} I_\alpha\right)$. For each $x \in F \cap \bigcup_{\alpha \ge 1} I_\alpha^\circ$, we can find arbitrarily small intervals $[y, z] \subset \bigcup_{\alpha \ge 1} I_\alpha^\circ$ such that $\frac{f(y)-f(z)}{y-z} \ge t$. We use Vitali's covering lemma to obtain a countable collection of disjoint intervals $\{J_\beta\}_{\beta \ge 1}$ such that

$$\bigcup_{\alpha \ge 1} I_\alpha^\circ \subset \bigcup_{\beta \ge 1} J_\beta$$

a.e. We also know $F \subset \bigcup_{\beta \ge 1} J_\beta$ a.e. Thus, we estimate

$$\lambda(F) \le \sum_{\beta \ge 1} \lambda(J_\beta) \le \frac{1}{t} \sum_{\beta \ge 1} \lambda\left(\widetilde{J}_\beta\right) \le \frac{1}{t} \lambda\left(\bigcup_{\alpha \ge 1} \widetilde{I}_\alpha\right) \le \frac{s}{t}(\lambda(F) + \epsilon)$$

Since $\epsilon > 0$ is arbitrary, we conclude that $\lambda(F) \le \frac{s}{t}\lambda(F)$. This complete the proof. □

Theorem 5.9. *If F is of bounded variation on $[a, b]$, then F' is integrable and*

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

PROOF. May assume that F is increasing. By Lebesgue's theorem, F' exists a.e. We set $F(x) = F(b)$ for $x > b$. Thus,

$$F'(x) = \lim_{k\to\infty} k[F(x+1/k) - F(x)] \qquad \text{for a.e. } x.$$

$$\begin{aligned}\int_a^b k[F(x+1/k) - F(x)]\,dx &= k\int_{a+1/k}^{b+1/k} F(x)\,dx - k\int_a^b F(x)\,dx \\ &= k\int_b^{b+1/k} F(x)\,dx - k\int_a^{a+1/k} F(x)\,dx \\ &= F(b) - k\int_a^{a+1/k} F(x)\,dx \le F(b) - F(a).\end{aligned}$$

Using Fatou's lemma, we have

$$\int_a^b F'(x)\,dx \le \liminf_{k\to\infty}\int_a^b k[F(x+1/k) - F(x)]\,dx \le F(b) - F(a).$$

□

EXAMPLE 5.7. If an increasing function F has a discontinuity at x, then there is a jump i.e. $F(x-) < F(x+)$. In this case, we have the strict inequality $F(b) - F(a) < \int_a^b F'(y)\,dy$, since F' do not count its jump. Hence, the continuity is a necessary condition to have FTC. Even if F is continuous and of bounded variation, it may not satisfy the Fundamental Theorem of Calculus. Recall the Cantor-Lebesgue function that is increasing, uniformly continuous, $F'(x) = 0$ a.e., but $F(1) = 1$, $F(0) = 0$.

Theorem 5.10. *If f is increasing on $[a, b]$, then f is continuous except for countably many points.*

PROOF. Exercise.

□

EXAMPLE 5.8. There is a monotone function which is discontinous at countably many points. Let $f(x) = 0$ when $x < 0$, and $f(x) = 1$ when $x \ge 0$. Choose a countable dense sequence $\{r_n\}$ in $[0, 1]$. Then,

$$F(x) = \sum_{n=1}^{\infty} \frac{1}{n^2} f(x - r_n)$$

is discontinuous at all points of the sequence $\{r_n\}$.

DEFINITION 5.4. An elementary jump function is a function $\sigma : \mathbb{R} \to \mathbb{R}$ which has the form:

$$\sigma(x)\begin{cases} a, & x < x_0 \\ b, & x = x_0 \\ c, & x > x_0 \end{cases}$$

for $a \le b \le c$. A function which can be written as a countable sum of elementary jump functions is called a jump function. i.e. $j(x) = \sum_{k=1}^{\infty} \sigma_k(x)$. jump function

Theorem 5.11. *(First decomposition) Let $F : [a, b] \to \mathbb{R}$ be a function of bounded variation. Then, F is decomposed into a jump function and a continuous function:*

$$F(x) = j(x) + g(x).$$

5.2.2. Absolute Continuity.

DEFINITION 5.5. We say $F : [a, b] \to \mathbb{R}$ is absolutely continuous if for any $\epsilon > 0$, there exists $\delta > 0$ such that for any finite union of disjoint intervals $\bigcup_{k=1}^{n}(x_k, y_k)$,

$$\sum_{k=1}^{n} |y_k - x_k| \leq \delta \quad \Rightarrow \quad \sum_{k=1}^{n} |F(y_k) - F(x_k)| \leq \epsilon.$$

REMARK 5.9.

(1) By dfinition, absolute continuity is stronger that uniform continuity, but weaker than Lipschitz continuity(i.e. $|F(x) - F(y)| \leq C|x - y|$ for some C.) In summary, $C^1 \Rightarrow$ Lipschitz continuity $\Rightarrow$ absolute continuity $\Rightarrow$ uniform continuity $\Rightarrow$ continuity.
(2) Absolutely continuous functions are of bounded variation. Decompose interval $[a, b]$ into small intervals of length δ. Then on each small interval $[x_k, x_{k+1}]$, we have $T_F(x_k, x_{k+1}) \leq \epsilon$. As the number of intervals is $(b - a)/\delta$, $T_F(a, b) \leq \frac{\epsilon(b-a)}{\delta}$.

EXAMPLE 5.10. There are functions which is uniformly continuous but not absolutely continuous.

$$f(x) = \begin{cases} x \sin \frac{1}{x}, & 0 < x \leq 1 \\ 0, & x = 0. \end{cases}$$

Theorem 5.12. *Let $F : [a, b] \to \mathbb{R}$ be measurable. Then F is absolutely continuous if and only if there is an integrable function f such that*

$$F(x) = F(a) + \int_a^x f(y)\, dy.$$

In this case, from Lebesgue differentiation theorem, $F' = f$ a.e.

Lemma 5.13. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a integrable function. Then, for any $\epsilon > 0$, there exists $\delta > 0$ such that for any $E \subset \mathbb{R}^n$ satisfying $\lambda(E) < \delta$,*

$$\int_E |f| < \epsilon.$$

PROOF. We may assume $f \geq 0$ without loss of generality. If $f \in L^\infty$, then the conclusion is obvious by choosing $\delta = \epsilon/N$. Now, we approximate f by f_N where $f_N(x) = \min(f(x), N)$ such that $\|f - f_N\|_1 < \epsilon/2$. (This is justified by MCT.) Then, we choose $\delta = \epsilon/(2N)$

$$\int_E |f| \leq \int_E |f - f_N| + \int_E |f_N| \leq \epsilon/2 + \epsilon/2N \cdot N = \epsilon.$$

□

Lemma 5.14. *If $F : [a, b] \to \mathbb{R}$ is absolutely continuous and $F' = 0$ a.e., then F is constant.*

Proof. Fix $c \in [a, b]$. We also fix small numbers $\epsilon_1, \epsilon_2 > 0$. It suffices to show

$$|F(c) - F(a)| \leq \epsilon_1 + \epsilon_2(c - a).$$

For $x \in E = \{y \in [a, c] : F'(y) = 0\}$, we can find arbitrarily small intervals $x \in [y, z]$ such that $\frac{F(z)-F(y)}{y-z} \leq \epsilon_2$. Collecting all such intervals for each $x \in E$, we can use Vitali covering's lemma to obtain a finite disjoint sub collection $I_1 = [y_1, z_1], \cdots, I_N = [y_N, z_N]$ such that $\lambda\left([a, c] \setminus \bigcup_{k=1}^N I_k\right) \leq \delta$, where δ satisfies the absolute continuity condition with respect to ϵ_1. We estimate

$$\begin{aligned}
|F(c) - F(a)| &\leq |F(a) - F(y_1)| + |F(y_1) - F(z_1) + |F(z_1) - F(y_2)| + \cdots |F(z_N) - F(c)| \\
&\leq \sum_{k=1}^{N} |F(y_k) - F(z_k)| + \sum_{k=1}^{N-1} |F(z_k) - F(y_{k+1})| + |F(a) - F(y_1)| + |F(z_N) - F(c)| \\
&\epsilon_2 \sum_{k=1}^{N} |y_k - z_k| + \epsilon_1 \leq \epsilon_2(c - a) + \epsilon_1
\end{aligned}$$

□

Proof. of *Theorem 5.12*
Lemma 5.13 shows 'if' part. We are left to show 'only if' part. Since F is absolutely continuous, it is of bounded variation and F' is integrable. We can define $G(x) = F(x) + \int_a^x F'(y)\, dy$. Then, we have $G'(x) = F'(x)$ a.e. by Lebesgue differentiation theorem. Hence, $G - F$ is absolute continuous and $(G - F)' = 0$ a.e. By Lemma 5.14, and from $F(a) = G(a)$, we conclude $G(x) = F(x)$. □

Finally, we summarize what we have done in this section in the following Theorem.

Theorem 5.15. *Lebesgue decomposition theorem Let $F : [a, b] \to \mathbb{R}$ be an increasing function. We have the following decomposition.*

$$F = j(x) + g(x) + \int_a^x f(y)\, dy,$$

where j is an increasing jump function, g is an increasing function with $g' = 0$ a.e., and f is a nonnegative integrable function.

APPENDIX A

Proof of Theorem 3.22

We do some preparation before giving a proof of Theorem 3.22

Lemma A.1. *Let $\mathcal{S}$ be the collection of those subsets of $X \times Y$ which are finite disjoint unions of measurable rectangles. Then $\mathcal{S}$ is an algebra.*

Proof. It is evident that the intersection of two measurable rectangles is a measurable rectangle. So, if $E, F \in \mathcal{S}$ then $E \bigcap F \in \mathcal{S}$. We are left to check the complement. Any measurable rectangle $A \times B$ is written as a union of two disjoint measurable rectangles,

$$(A \times B)^c = A^c \times Y \bigcup A \times B^c$$

. Thus, $(A \times B)^c \in \mathcal{S}$. If $E \in \mathcal{S}$, then $E = \bigcup_{j=1}^N E_j$, where E_j are measurable rectangles. Then, $E^c = \bigcap_{j+1}^N E_j^c$. Since $E_j^c \in \mathcal{S}$ and $\mathcal{S}$ is closed under finite intersection, $E^c \in \mathcal{S}$. Therefore, $\mathcal{S}$ is closed under finite union. $\square$

Definition A.1. Let X be a set and $\mathcal{S} \subset \mathcal{P}(X)$. Then $\mathcal{S}$ is a monotone class if $\mathcal{S}$ is closed under countable increasing unions and countable decreasing intersections. That is, if $A_j \in \mathcal{S}$ for $j = 1, 2, \cdots$, then

$$A_1 \subset A_2 \subset A_3 \subset \cdots, \quad \Rightarrow \quad \bigcup_{j=1}^{\infty} A_j \in \mathcal{S}$$
$$A_1 \supset A_2 \supset A_3 \supset \cdots, \quad \Rightarrow \quad \bigcap_{j=1}^{\infty} A_j \in \mathcal{S}$$

Any σ-algebra is a monotone class. Any intersection of monotone classes is a monotone class. For a collection $\mathcal{S}$ of subsets of X, we denote by $\mathcal{S}_m$ the intersection of all monotone classes containing $\mathcal{S}$, which is referred as the smallest monotone class containing $\mathcal{S}$ or a monotone class generated by $\mathcal{S}$.

Lemma A.2. *Let X be a set and $\mathcal{S} \subset \mathcal{P}(X)$ be an algebra. Then*

$$\mathcal{S}_m = \sigma(\mathcal{S}) =: \textit{the smallest } \sigma\textit{-algebra generated by } \mathcal{S}$$

Proof. Clearly, $\mathcal{S}_m \subset \sigma(\mathcal{S})$. In view of Lemma A.1, we are left to show $\mathcal{S}_m$ is an algebra. Let $A, B \in \mathcal{S}$.

- Claim: $A^c \in \mathcal{S}_m$
  Define $\mathcal{T} = \{A \subset X : A^c \in \mathcal{S}_m\}$. Then $\mathcal{S} \subset \mathcal{T}$. Moreover, $\mathcal{T}$ is a monotone class since

$\mathcal{S}_m$ is a monotone class. Indeed, for a increasing sequence $\{A_j\} \subset \mathcal{T}$, $A_j^c \in \mathcal{S}_m$ and so $\bigcap A_j^c \in \mathcal{S}_m$, $\bigcup A_j \in \mathcal{T}$. It is similar for a decreasing sequence. Therefore, $\mathcal{S}_m \subset \mathcal{T}$ and so for any $A \in \mathcal{S}_m$, we have $A^c \in \mathcal{S}_m$.

- Claim: $A \cup B \in \mathcal{S}_m$
  Fix $A \in \mathcal{S}$. Define $\mathcal{U} = \{B \subset X : A \cup B \in \mathcal{S}_m\}$. Then $\mathcal{S} \subset \mathcal{U}$ and $\mathcal{U}$ is a monotone class. It follows that $\mathcal{S}_m \subset \mathcal{U}$, that is, for any $B \in \mathcal{S}_m$ and $A \in \mathcal{S}$, we have $A \cup B \in \mathcal{S}_m$.
  Fix $B \in \mathcal{S}_m$ and define $\mathcal{V} = \{A \subset X : A \cup B \in \mathcal{S}_m\}$. Then by the same reasoning, we have $\mathcal{S}_m \subset \mathcal{V}$ and for any $A, B \in \mathcal{S}_m$, $A \cup B \in \mathcal{S}_m$.

□

Proof (Proof of Theorem 3.22).
Let $\mathcal{W}$ be the collection of all sets $E \in \mathcal{M} \times \mathcal{N}$ satisfying the theorem. We will show $\mathcal{W} = \mathcal{M} \times \mathcal{N}$. For that purpose, we want to show $\mathcal{W}$ is a σ-algebra. First of all, one can check that

$$\mathcal{W} \supset \mathcal{S} := \text{the algebra generated by } \{A \times B \subset X \times Y : A \in \mathcal{M}, B \in \mathcal{N}\}.$$

(Exercise)
In view of Lemma A.2, we are going to show $\mathcal{W}$ is a monotone class.

Claim: $\mathcal{W}$ is closed under countable disjoint unions.

Proof (Proof of Claim).
Let $\{E_j\}_{j=1}^{\infty} \subset \mathcal{W}$ be disjoint and $E = \cup_{j=1}^{\infty}$. Fix $y \in Y$. Then $E_y = \cup_{j=1}^{\infty} E_{k,y}$, which is a disjoint union. Thus, $E_y \in \mathcal{M}$. From the countable additivity of μ, $\mu(E_y) = \sum_{j=1}^{\infty} \mu(E_{k,y})$. Since $\mu(E_{j,y})$ are ν-measurable, $\mu(E_y)$ is also ν-measurable. Similarly, $E_x \in \mathcal{N}$ and $\nu(E_x)$ is μ-measurable.

$$\begin{aligned}
\int_Y \mu(E_y)\, d\nu(y) = \int_Y \sum_{j=1}^{\infty} \mu(E_{j,y})\, d\nu &= \sum_{j=1}^{\infty} \int_Y \mu(E_{j,y})\, d\nu && (\because \text{MCT}) \\
&= \sum_{j=1}^{\infty} \int_X \nu(E_{j,x})\, d\mu && (\because E_j \in \mathcal{W}) \\
&= \int_X \nu(E_x)\, d\mu
\end{aligned}$$

Therefore, $E \in \mathcal{W}$. □

Since $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ are σ-finite, we can write $X = \bigcup_{j=1}^{\infty} A_j$, $Y = \bigcup_{k=1}^{\infty} B_k$ where A_j, B_k are disjoint and $\mu(A_j), \nu(B_k) < \infty$. In order to show that $\mathcal{W}$ is a monotone class, bring a decreasing sequence $\{E_n\} \subset \mathcal{W}$. Then $E_n^{j,k} = E_n \bigcap (A_j \times B_k) \in \mathcal{W}$ for each j, k, n and $\{E_n^{j,k}\}$ is a decreasing sequence in n. We will show $E^{j,k} = \bigcap E_n^{j,k} \in \mathcal{W}$. Then $E = \bigcap E_n = \bigcup_{j,k} E^{j,k} \in \mathcal{W}$, thanks to Claim.
Fix j, k and we consider a decreasing sequence $\{E_n^{j,k}\}$. We omit the superscript for simplicity. For a fixed $y \in Y$, $E_y = \bigcap E_{n,y}$ ans so E_y is μ-measurable. Since $\mu(E_y) \le \mu(A_j) < \infty$, we have $\mu(E_y) = \lim_{n\to\infty} \mu(E_{n,y}) < \infty$. Thus, $y \mapsto \mu(E_y)$ is a ν-measurable function. Similarly, we have

E_x is ν-measurable and $x \mapsto \nu(E_x)$ is a μ-measurable function.

$$\begin{aligned}\int_Y \mu(E_y)\,d\nu(y) &= \lim_{n\to\infty}\int_Y \mu(E_{n,y}\,d\nu(y) \qquad (\because \text{MCT})\\ &= \lim_{n\to\infty}\int_X \nu(E_{n,x})\,d\mu(x)\\ &= \int_X \nu(E_x)\,d\mu(x)\end{aligned}$$

The proof for the increasing sequence is similar, in fact, even easier since one do not use the finiteness of measure. Therefore, we conclude that $\mathcal{W}$ is a monotone class and then by Lemma A.2 $\mathcal{W} = \mathcal{M}\times\mathcal{N}$.

We define a product measure $\pi(E)$ by the common value of (3). We remain to show the countable additivity. Let $E = \bigcup_{j=1}^{\infty} E_j$ be a countable disjoint union. Then $\{E_{j,y}\}$ are disjoint and,

$$\begin{aligned}\pi(E) = \int_Y \mu(E_y)\,d\nu(y) &= \int_Y \mu(\bigcup E_{j,y})\,d\nu(y)\\ &= \int_Y \sum_{j=1}^{\infty}\mu(E_{j,y})\,d\nu(y) = \sum_{j=1}^{\infty}\int_Y \mu(E_{j,y})\,d\nu(y)\\ &= \sum_{j=1}^{\infty}\pi(E_j).\end{aligned}$$

Finally, we show the uniqueness of such measures. We show for finite measure space, first. Assume that there two measures π_1, π_2 on $\mathcal{M}\times\mathcal{N}$, which agree on $\mathcal{S}$= the collection of finite unions of disjoint measurable rectangles. Define

$$\mathcal{T} = \{E \in \mathcal{M}\times\mathcal{N} : \pi_1(E) = \pi_2(E)\}.$$

Then, clearly, $\mathcal{S}\subset\mathcal{T}$. One can show $\mathcal{T}$ is a monotone class. (When you argue with a decreasing sequence, you need the finiteness of measure) Hence, $\mathcal{T} = \mathcal{M}\times\mathcal{N}$.

For σ-finite case, setting as before and fixing $A_j \times B_k$, we define a new measure $\widetilde{pi}_i(E) = \pi_i(E\cap (A_j\times B_k))$ for $i = 1, 2$. Then for any measurable rectangle $A\times B$, one can check $\widetilde{pi}_1(A\times B) = \widetilde{pi}_2(A\times B)$. Thus, by the finite measure case, $\widetilde{\pi}_1 = \widetilde{\pi}_2$ on $\mathcal{M}\times\mathcal{N}$. That is, for any $E\in\mathcal{M}\times\mathcal{N}$, $\pi_1(E\cap(A_j\times B_k)) = \pi_2(E\cap(A_j\times B_k))$. Then by countable additivity, we conclude that

$$\pi_1(E)\sum_{j,k}\pi_1(E\cap(A_j\times B_k)) = \sum_{j,k}\pi_2(E\cap(A_j\times B_k)) = \pi_2(E).$$

$\square$

Bibliography

[Fol] G. B. Folland, *Real analysis : mordern techniques and their application*, 2nd edition, Wiley.

[Jon] F. Jones, *Lebesgue integration on Euclidean space*, revised edition, Jones and Bartlett.

[Roy] H. L. Royden, *Real analysis*, 3rd edition, Prentice Hall.

[StSh] E. M. Stein, R. Shakarchi, *Real analysis*, Princeton Lecture Series in Analysis 3, Princeton Univercity Press.

[Tao] T. Tao, *An epsilon of room 1: pages from year three of a mathematical blog*, GSM **117**, AMS.