5-2023

BMO on Shapes

Benjamin Lewis Jeffers
Trinity University, benljeffers@gmail.com

BMO on Shapes

Benjamin L Jeffers

April 13, 2023

Contents

Chapter 1

Introduction

In this thesis we prove that the space BMO on shapes introduced by Dafni and Gibara is the dual space of the Hardy Space on Shapes which we introduce here for the first time. This document is relatively self contained, building up the necessary background in measure theory, functional, and harmonic analysis.

Chapter 2

Riemann Integral

There are many equivalent ways to define both the Riemann and the Lebesgue integral. We will introduce the Riemann integral using upper and lower piecewise constant integrals, and then define Lebesgue measure which will take us to the Lebesgue integral. I will also compare the Jordan measure, which can also be used to define the Riemann integral, with the Lebesgue measure to see some of the differences between the two. We will prove that if a function is Riemann integrable, then it is Lebesgue integral and the two are equal. This section will end by looking at a few examples of functions that are Lebesgue integrable but not Riemann integral. Most of this follows [Tao15a], [Tao15b], [Tao11].

Definition 1 (Partition)**.** We give two definitions of a partition. Let $I = [a, b], (a, b], [a, b)$, or (a, b) be an interval. Then a partition of I is a collection of increasing points in I,

$$\mathbb{P} = \{x_1, x_2, \ldots, x_n \mid x_i \in I, \text{ and } x_i < x_{i+1} \text{ for all } I\}.$$

More generally, given a set A, a partition of A is a collection of subsets $B_1, \ldots, B_n$ such that $B_i \cap B_j = \emptyset$ for all $i \neq j$, and $A = \bigcup_{i=1}^{n} B_i$.

Example. Consider the interval $I = [0, 3]$. A partition of I can be given as $\mathbb{P} = \left\{0, 1, \frac{3}{2}, 2, 3\right\}$, or equivalently, it could be given as $\mathbb{P} = \left\{[0, 1], (1, \frac{3}{2}), [\frac{3}{2}, 2], (2, 3]\right\}$, see figure 2.1.

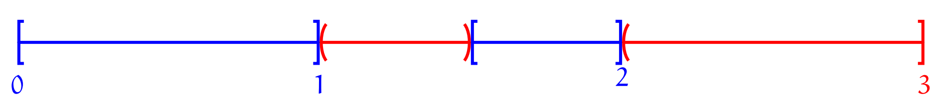


Figure 2.1: The partition $\mathbb{P} = \left\{[0, 1], (1, \frac{3}{2}), [\frac{3}{2}, 2], (2, 3]\right\}$

A non-example of a partition of $[0, 3]$ is $\mathbb{P}' = \{[0, 1], [2, 3]\}$. This is because it leaves out the part of the interval, $(1, 2)$ so it does not cover. Another non-example is $\mathbb{P}'' = \left\{[0, 1], (\frac{3}{4}, \frac{3}{2})[\frac{3}{2}, 2), [2, 3]\right\}$. This one is not a partition because two elements of the partition overlap $[0, 1] \cap (\frac{3}{4}, \frac{3}{2}) \neq \emptyset$. These can be seen in figure 2.2.

Definition 2 (Piecewise Constant Function)**.** Let I be a bounded interval, $f\colon I \to \mathbb{R}$ be a function, and $\mathbb{P}$ be a partition of I. We say that f is *piecewise constant with respect to* $\mathbb{P}$ if for every $J \in \mathbb{P}$, f is constant on J.

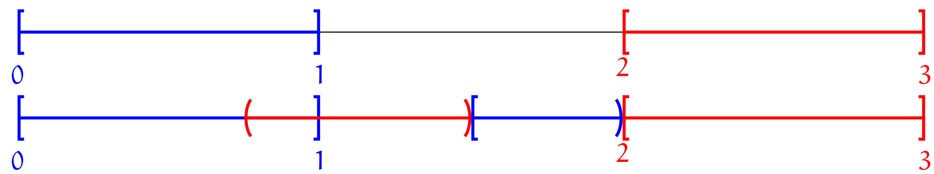


Figure 2.2: The non-partitions $\mathbb{P}' = \{[0,1],[2,3]\}$ and $\mathbb{P}'' = \left\{[0,1],(\frac{3}{4},\frac{3}{2})[\frac{3}{2},2),[2,3]\right\}$.

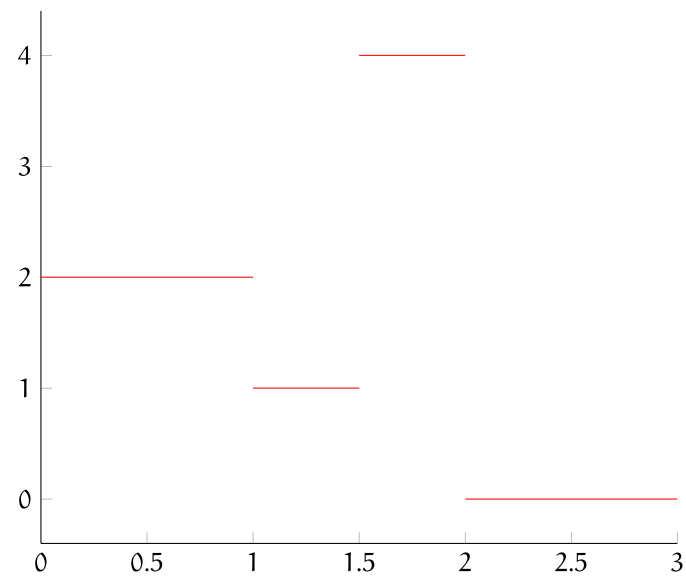


Figure 2.3: Graph of the piecewise constant function f.

Example. An example of such a function on our partition from above is

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x < \frac{3}{2} \\ 4 & \text{if } \frac{3}{2} \le x \le 2 \\ 0 & \text{if } 2 < x \le 3. \end{cases}$$

Using this definition of piecewise constant functions allows us to define what their integral should be from our intuition of area.

Definition 3 (Piecewise Constant Integral for a Partition)**.** Let I be a bounded interval, let $\mathbb{P}$ be a partition of I. Let $f\colon I \to \mathbb{R}$ be a function which is piecewise constant with respect to $\mathbb{P}$. Then we define the *piecewise constant integral* p.c. $\int_{[\mathbb{P}]} f$ of f with respect to the partition $\mathbb{P}$ by the formula

$$\text{p.c.} \int_{[\mathbb{P}]} f := \sum_{J \in \mathbb{P}} c_J \, |J| \, ,$$

where for each J in $\mathbb{P}$, we let c_J be the constant value of f on J, and $|J|$ be the length of the interval.

Example. When we take the integral of the function $f\colon [0,3] \to \mathbb{R}$ defined as above, we calculate the area under each constant piece. On the first part of the partition $[0,1]$, the value that f takes is 2, so $2 \cdot |I_1| = 2 \cdot 1 = 2$, will be the first summand in the integral. Similarly, for the next interval, it takes the value 1 from $[1, 1.5]$, so $1 \cdot |I_2| = 1 \cdot \frac{1}{2} = \frac{1}{2}$. We continue in this way for each element of the partition to obtain

$$\text{p.c.} \int_{[\mathbb{P}]} f = 1 \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 4 + 0 \cdot 1 = \frac{9}{2}.$$

We prove a quick preliminary lemma that will be helpful for the upcoming proposition.

Theorem 4 (Length is Finitely Additive, [Tao15a] Theorem 11.1.13)**.** *Let* I *be a bounded interval,* n *be a natural number, and let* $\mathbb{P}$ *be a partition of* I *of cardinality* n*. Then*

$$|I| = \sum_{J \in \mathbb{P}} |J|.$$

Proof. We proceed by induction on n. More precisely, let $P(n)$ be the property that whenever I is a bounded interval, and whenever $\mathbb{P}$ is a partition of I with cardinality n, then $|I| = \sum_{J \in \mathbb{P}} |J|$.

The base case, $P(0)$, just says that the only way I can be partitioned into an empty partition is if I itself is empty. Similarly, the case $P(1)$ follows from the fact that there is only one way to partition I into one piece, so $I = J$.

Now, suppose that $P(n)$ is true for some $n \geq 1$, and now we prove $P(n+1)$. Let I be a bounded interval and let $\mathbb{P}$ be a partition of I of cardinality $n+1$. If I is either the empty set or a point, then all the intervals in $\mathbb{P}$ must also be either the empty set or a point, and so they both have length zero, and the claim follows. So we may assume that I is an interval of the form $(a, b), [a, b), (a, b]$, or $[a, b]$.

Let us first suppose that $b \in I$. Since $b \in I$, we know that one of the intervals K in $\mathbb{P}$ contains b. Since K is contained in I, it must therefore be of the form $(c, b], [c, b]$, or $\{b\}$ for some real number c with $a \leq c \leq b$. In particular, this means that the set $I - K$ is also an interval of the form $[a, c], (a, c), (a, c], [a, c)$ when $c > a$, or a point or empty set when $a = c$. Either way, we can see that

$$|I| = |K| + |I - K|.$$

On the other hand, since $\mathbb{P}$ form a partition of I, we see that $\mathbb{P} - K$ forms a partition of $I - K$. By the induction hypothesis, we thus have

$$|I - K| = \sum_{J \in \mathbb{P} - \{K\}} |J|.$$

Combining these two identities we obtain

$$|I| = \sum_{J \in \mathbb{P}} |J|$$

as desired.

Now suppose that $b \notin I$, i.e., I is either (a, b) or $[a, b)$. Then one of the intervals K is also of the form (c, b) or $[c, b)$. In particular, this means that the set $I - K$ is also an interval of the form $[a, c], (a, c), (a, c], [a, c)$ when $c > a$ or a point or empty set when $a = c$. The rest of the argument then proceeds as above. □

Now the natural question to ask is whether or not the definition of this integral depends on the chosen partition. The following proposition tells us that it does not.

Proposition 5 ([Tao15a] Proposition 11.2.13)**.** *Let* I *be a bounded interval, and let* $f\colon I \to \mathbb{R}$ *be a function. Suppose that* $\mathbb{P}$ *and* $\mathbb{P}'$ *are partitions of* I *such that* f *is piecewise constant both with respect to* $\mathbb{P}$ *with respect to* $\mathbb{P}'$*. Then*

$$\text{p.c.}\int_{[\mathbb{P}]} f = \text{p.c.}\int_{[\mathbb{P}']} f.$$

Proof. Consider the common refinement

$$\mathbb{P}\#\mathbb{P}' = \left\{K \cap J \mid K \in \mathbb{P} \text{ and } J \in \mathbb{P}'\right\}.$$

Now we want to show that

$$\text{p.c.}\int_{[\mathbb{P}]} f = \text{p.c.}\int_{[\mathbb{P}\#\mathbb{P}']} f = \text{p.c.}\int_{[\mathbb{P}']} f.$$

So we have

$$\text{p.c.}\int_{[\mathbb{P}]} f = \sum_{J\in\mathbb{P}} c_J\,|J|\,,\ \text{p.c.}\int_{[\mathbb{P}']} f = \sum_{K\in\mathbb{P}'} d_K\,|k|\,,\ \text{p.c.}\int_{[\mathbb{P}\#\mathbb{P}']} f = \sum_{L\in\mathbb{P}\#\mathbb{P}'} e_L\,|L|\,.$$

Since length is finitely additive by Theorem 4, for any $J \in \mathbb{P}$, there exist $J_1, \ldots, J_n \in \mathbb{P}\#\mathbb{P}'$ such that each $J_i \subseteq J$. Since $\mathbb{P}\#\mathbb{P}'$ is a partition, these are disjoint and $J_1 \cup \cdots \cup J_n = J$, and so

$$|J| = \sum_{i=1}^{n} |J_i|\,.$$

Indeed, let $x \in J$, then $x \in I$, and $\mathbb{P}'$ is a partition of I, so there exists a $K \in \mathbb{P}'$ such that $x \in K$. Then $x \in J \cap K = J_i$ for some i. Now suppose $x \in J_1 \cup \cdots \cup J_n$. Then $x \in J_i$ for some i, and $J_i = J' \cap K'$ for some $J' \in \mathbb{P}$ and $K' \in \mathbb{P}'$, but since $J' \subseteq J$, $J' = J$ since otherwise $\mathbb{P}$ would not be a partition. Thus they are equal. So we get that

$$\text{p.c.}\int_{[\mathbb{P}]} f = \sum_{J\in\mathbb{P}} c_J\,|J| = \sum_{L\in\mathbb{P}\#\mathbb{P}'} e_L\,|L| = \text{p.c.}\int_{[\mathbb{P}\#\mathbb{P}']} f.$$

An analogous result holds for $\mathbb{P}'$, and thus we get the desired statement. □

Example. For example, consider the function $f\colon [0,3] \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x < \frac{3}{2} \\ 4 & \text{if } \frac{3}{2} \le x \le 2 \\ 0 & \text{if } 2 < x \le 3. \end{cases}$$

If we change our partition to be $\mathbb{P}' = \left\{[0, \frac{3}{4}), [\frac{3}{4}, 1], (1, \frac{3}{2}), [\frac{3}{2}, 2], (2, \frac{9}{4}], (\frac{9}{4}, 3]\right\}$, then our function is still piecewise constant with respect to this partition:

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x < \frac{3}{4} \\ 2 & \text{if } \frac{3}{4} \le x \le 1 \\ 1 & \text{if } 1 < x < \frac{3}{2} \\ 4 & \text{if } \frac{3}{2} \le x \le 2 \\ 0 & \text{if } 2 < x \le \frac{9}{4} \\ 0 & \text{if } \frac{9}{4} < x \le 3. \end{cases}$$

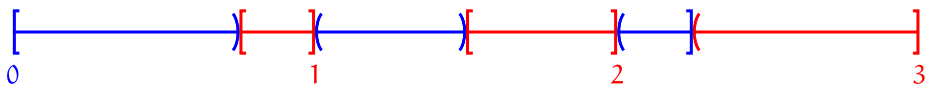


Figure 2.4: The partition $\mathbb{P}' = \left\{[0, \frac{3}{4}), [\frac{3}{4}, 1], (1, \frac{3}{2}), [\frac{3}{2}, 2], (2, \frac{9}{4}], (\frac{9}{4}, 3]\right\}$

Then calculating the integral of f with respect to both $\mathbb{P}$ and $\mathbb{P}'$ we get

$$\text{p.c.} \int_{[\mathbb{P}]} f = 1 \cdot 2 + \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 4 + 1 \cdot 0 = \frac{9}{2}.$$
$$\text{p.c.} \int_{[\mathbb{P}']} f = 2 \cdot \frac{3}{4} + 2 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} + 0 \cdot \frac{1}{4} + 0 \cdot \frac{3}{4} = \frac{9}{2}.$$

Due to this proposition we can define the piecewise constant integral over the interval I, instead of just for a particular partition.

Definition 6 (Piecewise Constant Integral for an Interval)**.** Let I be a bounded interval, and let $f\colon I \to \mathbb{R}$ be a piecewise constant function on I. We define the *piecewise constant integral* p.c. $\int_I f$ by the formula

$$\text{p.c.} \int_I f = \text{p.c.} \int_{[\mathbb{P}]} f,$$

where $\mathbb{P}$ is any partition of I with respect to which f is piecewise constant.

Example. In our running example of $f\colon [0, 3] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x < \frac{3}{2} \\ 4 & \text{if } \frac{3}{2} \le x \le 2 \\ 0 & \text{if } 2 < x \le 3. \end{cases}$$

we have p.c. $\int_I f = \frac{9}{2}$.

To proceed to define the Riemann integral of a general function. We need a few other definitions first.

Definition 7 (Majorize and Minorize)**.** Let $f\colon I \to \mathbb{R}$ and $g\colon I \to \mathbb{R}$. We say that g *majorizes* f on I if we have $g(x) \ge f(x)$ for all $\in I$, and that g *minorizes* f on I if $g(x) \le f(x)$ for all $x \in I$.

Example. For example, if $I = [0, 1]$, and $f(x) = x^2$, then $g(x) = x$ majorizes f since for all $x \in I$, $x^2 \le x$. Similarly, $h(x) = x^3$ minorizes x^2 since for all $x \in I$, $x^3 \le x^2$. In our running example of $f\colon [0, 3] \to \mathbb{R}$, defined by

$$f(x) = \begin{cases} 2 & \text{if } 0 \le x \le 1 \\ 1 & \text{if } 1 < x < \frac{3}{2} \\ 4 & \text{if } \frac{3}{2} \le x \le 2 \\ 0 & \text{if } 2 < x \le 3. \end{cases}$$

the function $g'(x) = 5$ majorizes f while the function $h(x) = -3$ minorizes it.

Definition 8 (Upper and Lower Riemann Integral)**.** Let $f\colon I \to \mathbb{R}$ be a bounded function defined on a bounded interval I. We define the *upper Riemann integral* $\overline{\int}_I f$ by the formula

$$\overline{\int}_I f = \inf\left\{\text{p.c.}\int_I g \mid g \text{ is a p.c. function on } I \text{ which majorizes } f\right\}$$

and the *lower Riemann integral* $\underline{\int}_I f$ by the formula

$$\underline{\int}_I f = \sup\left\{\text{p.c.}\int_I g \mid g \text{ is a p.c. function on } I \text{ which minorizes } f\right\}.$$

Example. Consider the function $f(x) = x$ on the interval $I = [0, 1]$. Looking at the definition of the upper Riemann integral, we need to bound this function from above by a piecewise constant function. As we take the infimum, the piecewise constant function approach our function as in figure 2.5.

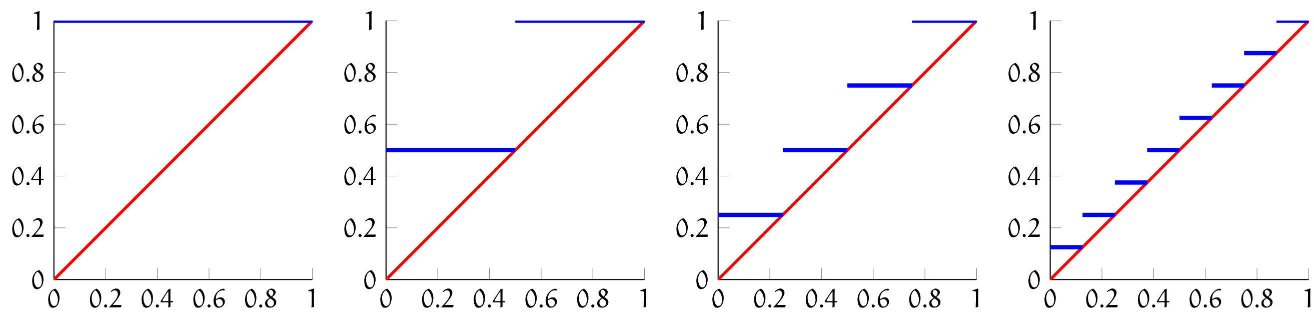


Figure 2.5: Approximation of $f(x) = x$ by piecewise constant functions from above.

We can do the same thing approximating from below as in figure 2.6,

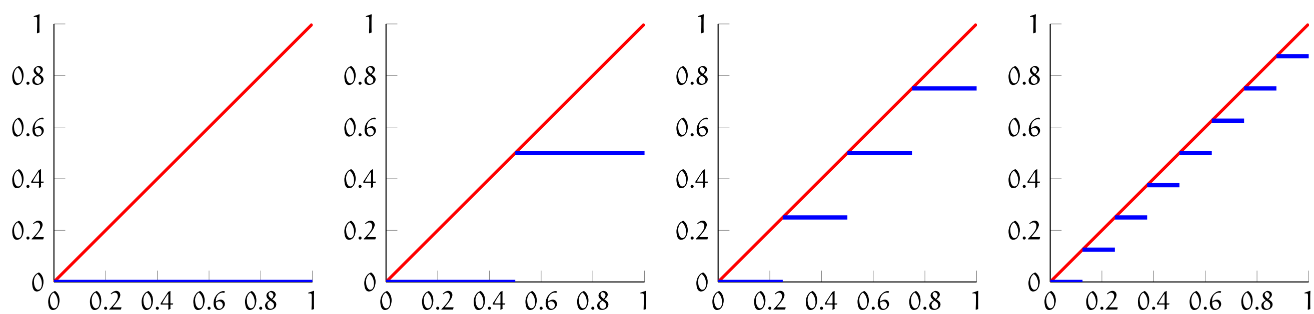


Figure 2.6: Approximation of $f(x) = x$ by piecewise constant functions from below.

Finally, the general Riemann integral is defined if the upper and lower Riemann integrals are equal.

Definition 9 (Riemann Integral). Let $f\colon I \to \mathbb{R}$ be a bounded function on a bounded interval I. If $\underline{\int}_I f = \overline{\int}_I f$, then we say that f is Riemann integrable on I and define

$$\int_I f = \underline{\int}_I f = \overline{\int}_I f.$$

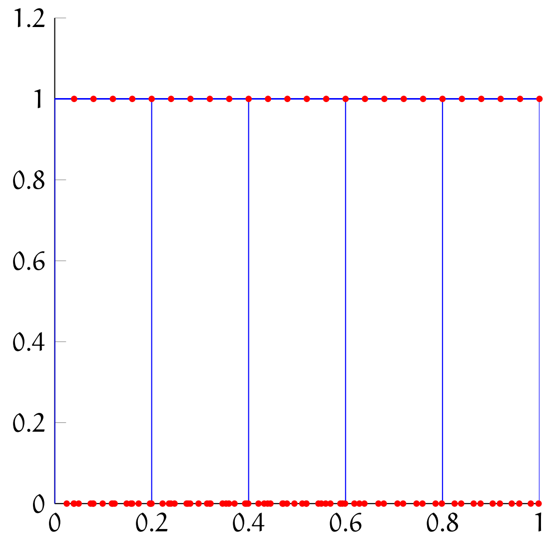


Figure 2.7: Upper Riemann sums of Dirichlet's function

Remark 10. While this definition is quite sufficient and worked for hundreds of years, it turns out to be insufficient in some ways as shown in the next example.

Example. Consider the function $f\colon [0,1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

This is called Dirichlet's function and is shown in figure 2.7. In this case, any piecewise constant function g that majorizes f must be such that $g(x) \geq 1$ except on a finite number of points. Suppose that $g(x) < 1$ on some interval $I \subset [0,1]$. Then unless I is a single point, there will be a rational number $r \in I$, so then $f(r) = 1$, but $g(r) < 1$, so g does not majorize f. So I must be a single point, however, when we calculate the integral, this point has length 0, so it will not contribute to the integral at all. Thus, it can be seen that that infimum of all piecewise constant functions that majorize f is the function $g(x) = 1$, the integral of this function over $[0,1]$ is

$$\text{p.c.} \int_I g(x)dx = 1 \cdot 1 = 1.$$

A similar argument shows that the supremum of all piecewise constant functions that minorize f is the function $h(x) = 0$, and we have

$$\text{p.c.} \int_I h(x)dx = 0 \cdot 1 = 0.$$

Since the upper and lower Riemann integrals do not agree, this function is not Riemann integrable.

Another way to define the Riemann integral is to use Riemann sums which we briefly review here because it will be easier to prove that certain functions are not Riemann integrable using this formulation.

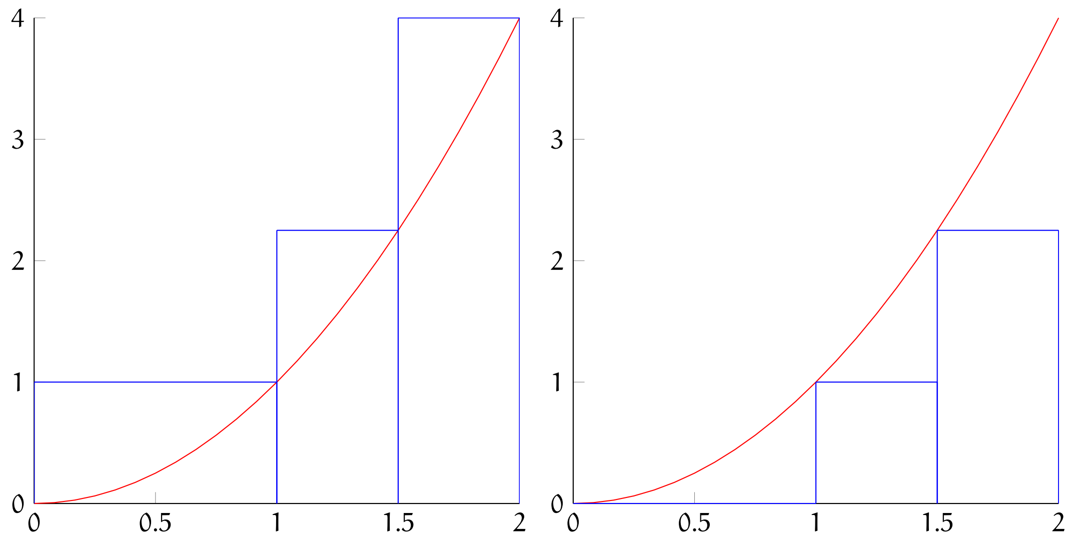


Figure 2.8: The upper Riemann sum $U(x^2, \{[0,1], (1,1.5], (1.5,2]\}$ and the lower Riemann sum $L(x^2, \{[0,1], (1,1.5], (1.5,2]\}$

Definition 11 (Riemann Sums)**.** Let $f\colon I \to \mathbb{R}$ be a bounded function on a bounded interval I, and let $\mathbb{P}$ be a partition of I. We define the *upper Riemann sum* $U(f, \mathbb{P})$ and the *lower Riemann sum* $L(f, \mathbb{P})$ by

$$U(f, \mathbb{P}) = \sum_{J \in \mathbb{P}, J \neq \emptyset} |J| \cdot \sup\{f(x) \mid x \in J\}$$

and

$$L(f, \mathbb{P}) = \sum_{J \in \mathbb{P}, J \neq \emptyset} |J| \cdot \inf\{f(x) \mid x \in J\}.$$

Example. Consider the function $f(x) = x^2$ on the interval $[0,2]$. We can partition the interval as $\mathbb{P} = \{[0,1], (1,1.5], (1.5,2]\}$, although any other partition works as well. We see that the supremum of x^2 on the interval $[0,1]$ is 1, while the supremums of x^2 on the intervals $(1,1.5]$ and $(1.5,2]$ are 2.25, and 4, respectively. So we have

$$U(x^2, \{[0,1], (1,1.5], (1.5,2]\} = 1 \cdot 1 + 2.25 \cdot \frac{1}{2} + 4 \cdot \frac{1}{2} = 4.125.$$

We know from our basic calculus classes that the integral should be $\int_0^2 x^2\,dx = \frac{1}{3}x^3|_0^2 = \frac{1}{3} \cdot 2^3 - \frac{1}{3}0^3 = \frac{8}{3} \approx 2.666$, so our estimate with this Riemann sum is fairly far off, this is because our partition still has pretty big intervals. As we make them narrower the Riemann sum will converge to the area under the curve. Similarly, we can calculate the lower Riemann sum of this function with respect to $\mathbb{P}$. In this case we see that the infimum of x^2 on the three intervals is 0, 1, and 2.25. respectively. Thus we have

$$U(x^2, \{[0,1], (1,1.5], (1.5,2]\} = 0 \cdot 1 + 1 \cdot \frac{1}{2} + 2.25 \cdot \frac{1}{2} = 1.625.$$

Both of these are illustrated in figure 2.8.

Then we have the following two propositions that connect the Riemann sums to the Riemann integral.

Lemma 12 ([Tao15a] Lemma 11.3.11)**.** *Let* $f\colon I \to \mathbb{R}$ *be a bounded function on a bounded interval* I *and let* g *be a function which majorizes* f *and which is piecewise constant with respect to some partition* $\mathbb{P}$ *of* I*. Then*

$$\text{p.c.}\int_I g \geq U(f, \mathbb{P}).$$

Similarly, if h *is a function which minorizes* f *and is piecewise constant with respect to* $\mathbb{P}$*, then*

$$\text{p.c.}\int_I h \leq L(f, \mathbb{P}).$$

Proof. Suppose that $\text{p.c.}\int_I g < U(f,\mathbb{P}) = \sum_{J\in\mathbb{P}} |J| \sup\{f(x) \mid x \in J\}$. Since $\text{p.c.}\int_I g = \sum_{J\in\mathbb{P}} c_J\,|J|$, and since $\mathbb{P}$ is the same in both cases, there must be some $J \in \mathbb{P}$ such that

$$c_J\,|J| < |J| \cdot \sup\{f(x) \mid x \in J\}.$$

But then this means that there is an $x \in J$ such that $c_J < f(x) < \sup\{f(x) \mid x \in J\}$, a contradiction to the fact that g majorizes f. Thus

$$\text{pc.}\int_I g \geq U(f, \mathbb{P}).$$

An analogous method can be used to show that

$$\text{p.c.}\int_I h \leq L(f, \mathbb{P})$$

if h minorizes f. ☐

Proposition 13 ([Tao15a] Proposition 11.3.12)**.** *Let* $f\colon I \to \mathbb{R}$ *be a bounded function on a bounded interval* I*. Then*

$$\overline{\int}_I f = \inf\{U(f,\mathbb{P}) \mid \mathbb{P} \textit{ is a partition of } I\}$$

and

$$\underline{\int}_I f = \sup\{L(f,\mathbb{P}) \mid \mathbb{P} \textit{ is a partition of } I\}.$$

Proof. From lemma 12, we can see immediately that $\overline{\int}_I f \geq \inf\{U(f,\mathbb{P}) \mid \mathbb{P} \text{ is a partition on } I\}$. For the other direction, suppose that

$$\overline{\int}_I f > \inf\{U(f,\mathbb{P}) \mid \mathbb{P} \text{ is a partition on } I\}.$$

Then there exists a partition $\mathbb{P}$ such that

$$\sum_{J\in\mathbb{P}} |J| \cdot \sup\{f(x) \mid x \in J\} = U(f,\mathbb{P}) < \overline{\int}_I f \leq \text{p.c.}\int_I g,$$

where g is any majorizing piecewise constant function. Since $\sup_{x\in J} \geq f(x)$ for all $x \in J$ it follows that the function h that takes the value $h(x) = \sup_{x\in J} f(x)$ on J is a majorizing function, and we obtain a contradiction

$$\sum_{J\in\mathbb{P}}^{n} |J| \cdot \sup\{f(x) \mid x \in J\} < \text{p.c.}\int_I h = \sum_{J\in\mathbb{P}} |J| \cdot \sup\{f(x) \mid x \in J\}.$$

An analogous method can be used to obtain the other half of the proposition. ☐

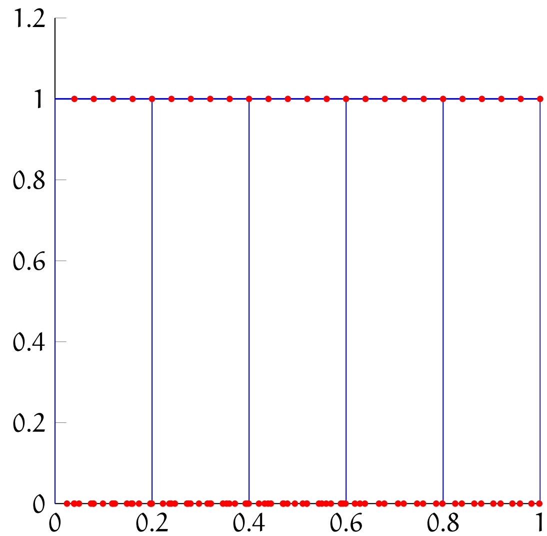


Figure 2.9: Upper Riemann sums of Dirichlet's function

We quickly show that the Riemann integral is insufficient for some functions. Consider Dirichlet's function $f\colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

This function is bounded, so we show that it is not Riemann integrable. Let $\mathbb{P}$ be any partition of $[0,1]$. If J is not a point or an empty set, then

$$\sup \{f(x) \mid x \in J\} = 1,$$

and in particular this gives

$$|J| \cdot \sup \{f(x) \mid x \in J\} = |J|.$$

Thus we have

$$U(f, \mathbb{P}) = \sum_{J \in \mathbb{P}, J \neq \emptyset} |J| = 1.$$

Therefore since the empty set does not contribute to length, we have that $\overline{\int}_{[0,1]} f = 1$. We can see that in figure 2.9 that no matter how we partition $[0,1]$, there will always be a rational number in each interval, and an irrational number in each interval.

Similarly, we have that

$$\inf \{f(x) \mid x \in J\} = 0$$

and thus

$$L(f, \mathbb{P}) = \sum_{J \in \mathbb{P}, J \neq \emptyset} 0 = 0.$$

So $\underline{\int}_{[0,1]} f = 0$. Thus since the upper and lower Riemann integral are not equal, the function is not Riemann integrable.

Intuitively, the integral of this function should be 0, since there are far less rational numbers than irrational numbers, so the nonzero part of the integral should not be able to contribute much. This issue will be fixed when we move to the Lebesgue integral.

Remark 14. Another reason the Riemann integral is insufficient is that if $\{f_n\}_{n=1}^{\infty}$ is a sequence of functions that tends to f pointwise on some interval I, as n tends to infinity, then we do not necessarily have

$$\lim_{n\to\infty} \int_I f_n \, dx = \int_I f dx.$$

In developing the Lebesgue measure, a goal will be to prove that this holds for Lebesgue integrals.

Finally we say a few words about Jordan measure, this was a reasonable first measure, but we will soon see that it is insufficient for a robust theory. First we might want to consider what a measure is and what it should do. There are any number of reasons why we might want to be able to measure the size of sets, the size of a set is one of its fundamental properties. There are many difficulties in defining this measure, since most sets we might want to measure are infinite. Intuitively, a point should have size 0, but if we want to measure a set by summing up its components, we have an infinite number of points, each with size 0, so we would have to make sense of $0 \cdot \infty$. There are other issues. The sets $[0, 1]$ and $[0, 2]$ are in bijection with each other, but a sensible definition of measure would give the interval $[0, 1]$, a length of 1, while the set $[0, 2]$ should have a length of 2. There are many ways of going about solving these problems, but there are three rules we would like a measure to satisfy. Calling our measure m, we should have

(a) $m(\emptyset) = 0$, the empty set should not have a positive measure.

(b) If $E \subset F$, then $m(E) \leq m(F)$, since F contains E it should be the same size or bigger.

(c) If $E_1, E_2, \dots$ are disjoint sets, then $m(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} m(E_i)$, consider the sets $E_i = [i, i+1]$, then the measure of the union of all of these sets should be infinite. For a finite example, consider the sets $E_i = [\frac{1}{2^i}, \frac{1}{2^{i-1}}]$, so $E_1 = [\frac{1}{2}, 1]$, $E_2 = [\frac{1}{4}, \frac{1}{2}], \dots$. We know from calculus that the series $\sum_{i=1}^{\infty} \frac{1}{2^n} = 1$, so we would want the measure of the union of these sets to be 1.

Definition 15 (Volume)**.** We define the *length* of an interval $I = [a, b], [a, b), (a, b], (a, b)$ to be $|I| = a - b$. A *box* in $\mathbb{R}^d$ is a Cartesian product $B = I_1 \times \cdots \times I_d$ of d intervals $I_1 \dots, I_d$. The *volume* $|B|$ of such a box is defined as

$$|B| = |I_1| \times |I_2| \times \cdots \times |I_d| .$$

An *elementary set* is an subset of $\mathbb{R}^d$ which is the union of a finite number of boxes.

Example. The volume of boxes in dimension 0, 1, 2, and 3, can be seen in figure 2.10.

An example of an elementary set can be seen in figure 2.11. Note that the boxes that make up the elementary set do not need to be of the same dimension.

Lemma 16 ([Tao11] Lemma 1.1.2)**.** *Let $E \subset \mathbb{R}^d$ be an elementary set.*

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

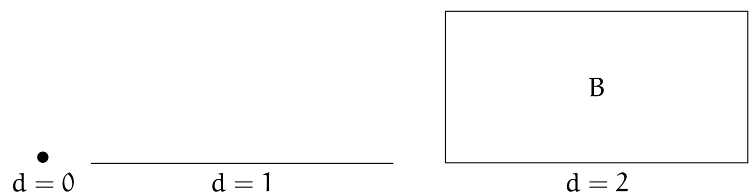


Figure 2.10: Volume of boxes in dimensions 0, 1, and 2.

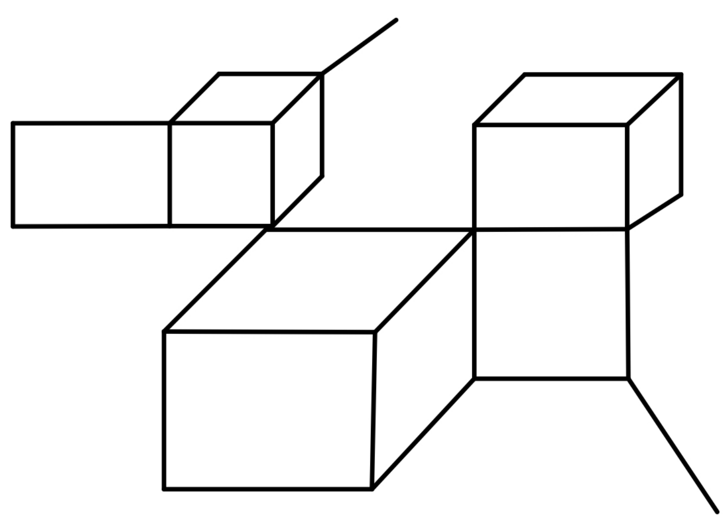


Figure 2.11: Example of an elementary set in $\mathbb{R}^3$

(b) If E *is partitioned as the finite union* $B_1 \cup \cdots \cup B_k$ *of disjoint boxes, then the quantity* $m(E) = |B_1| + \cdots + |B_k|$ *is independent of the partition.*

We refer to $m(E)$ as the elementary measure of E. The main idea of the proof of the second part of the above proposition is that the length of an interval can be written as $|I| = \lim_{N\to\infty} \frac{1}{N}\#(I \cap \frac{1}{N}\mathbb{Z})$. This works fairly well for elementary sets, but we do not want to define the measure of an arbitrary set $E \subseteq \mathbb{R}^d$ as

$$m(E) = \lim_{N\to\infty} \frac{1}{N^d}\#(E \cap \frac{1}{N}\mathbb{Z}^d).$$

This will not obey some basic properties that a measure should have. One of these is that if E has a measure A, then translating E by a vector $\vec{x}$, then $m(E + \vec{x}) = A$ as well. However, if we let $E = \mathbb{Q} \cap [0, 1]$, then this definition would give a measure of 1, but the translate $E + \sqrt{2}$ will have a measure of zero, since every element of this set will be irrational. So we need something more robust, especially since most of the sets we are interested in will not be elementary sets. One way to do this is with the *Jordan measure*.

Definition 17 (Jordan Measure)**.** Let $E \subseteq \mathbb{R}^d$ be a bounded set.

(a) The *Jordan inner measure* $m_{*,(J)}(E)$ of E is defined as

$$m_{*,(J)}(E) = \sup\{m(A) \mid A \subset E, A \text{ elementary}\}.$$

(b) The *Jordan outer measure* $m^{*,(J)}(E)$ is defined as

$$m^{*,(J)}(E) = \inf\{m(B) \mid B \supset E, B \text{ elementary}\}.$$

(c) If $m_{*,(J)} = m^{*,(J)}$, then we say that E is Jordan measurable and call $m(E) = m_{*,(J)} = m^{*,(J)}$ the *Jordan measure* of E.

Example. The simplest example of a Jordan measurable set is an elementary set E. In this case E is an elementary set containing E, and no smaller elementary set can contain E, so

$$\inf\{m(B) \mid B \supset E, B \text{ elementary}\} = m(E)$$

. Similarly, E is the largest elementary set that is contained in E, so it is also its own inner measure. Thus E is Jordan measurable, in particular any rectangle in $\mathbb{R}^2$ is Jordan measurable.

This measure satisfies several nice properties.

Proposition 18 ([Tao11] Exercise 1.1.6)**.** *Let* $E, F \subset \mathbb{R}^d$ *be Jordan measurable sets. Then the following hold:*

(a) $E \cup F, E \cap F, E \backslash F, E \Delta F = (E\backslash F) \cup (F\backslash E)$ *are Jordan measurable.*

(b) $m(E) > 0$.

(c) If E, F *are disjoint, then* $m(E \cup F) = m(E) + m(F)$.

(d) If $E \subset F$*, then* $m(E) \leq m(F)$.

(e) $m(E \cup F) \leq m(E) + m(F)$.

(f) *For any* $x \in \mathbb{R}^d$, $E + x$ *is Jordan measurable, and* $m(E + x) = m(E)$.

The reason we mention the Jordan measure here is because we can reformulate the Riemann integral in terms of the Jordan measure.

Proposition 19 ([Tao11] Exercise 1.1.25)**.** *Let* $[a, b]$ *be an interval, and let* $f\colon [a, b] \to \mathbb{R}$ *be a bounded function. Then* f *is Riemann integrable if and only if the sets* $E_+ = \{(x, t) \mid x \in [a, b], 0 \leq t \leq f(x)\}$ *and* E_- *are both Jordan measurable in* $\mathbb{R}^2$*, in which case on has*

$$\int_a^b f(x)dx = m(E_+) - m(E_-).$$

This definition gives us the interpretation of the integral as the area under the curve.

While the Jordan measure is more useful than the elementary measure, there are still many places where the Jordan measure might not be sufficient. One example is on the fat Cantor set. This set is constructed by first taking the interval $[0, 1]$ and removing the middle $\frac{1}{4}$ leaving us with $[0, \frac{3}{8}] \cup [\frac{5}{8}, 1]$. Then we proceed by removing the middle $\frac{1}{4^n}$ out of each of the remaining 2^{n-1} intervals. This has Jordan outer measure 1 since for any covering of this set with finite boxes, each box must contain a point of the set. However, it has Jordan inner measure 0 since it contains no intervals. Thus it is not Jordan measurable. Our next task is to investigate the Lebesgue measure. It is able to measure these kinds of sets that the Jordan measure cannot, and it also gives us a completion of the Riemann integral.

Chapter 3

Lebesgue Measure

To get a more robust measure we first start by changing the definition of the Jordan outer measure. When we defined the Jordan outer measure of a set A we took the infimum of the measure of all elementary sets that contained A. The elementary sets were finite unions of boxes. We will get a much better measure by changing finite to countable, and looking at countable unions of boxes that contain our set.

Definition 20 (Lebesgue Outer Measure)**.** We define the *Lebesgue outer measure* $m^*(E)$ of E as

$$m^*(E) = \inf \left\{ \sum_{n=1}^{\infty} |B_n| \mid \bigcup_{n=1}^{\infty} B_n \supset E; B_1, \dots, \text{ boxes} \right\}.$$

Theorem 21 ([Tao11] Exercise 1.2.3)**.** *The Lebesgue outer measure* m^* *satisfies the following:*

(a) (Empty set) $m^*(\emptyset) = 0$.

(b) (Monotonicity) If $E \subset F \subset \mathbb{R}^d$*, then* $m^*(E) \leq m^*(F)$.

(c) (Countable subadditivity) If $E_1, E_2, \dots \subset \mathbb{R}^d$ *is a countable sequence of sets, then*

$$m^* \left(\bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n=1}^{\infty} m^*(E_n)$$

, and there is equality in the case where the E_i *are pairwise disjoint.*

Proof. Properties (a) and (b) come straight from the definition of Lebesgue measure. For property (c) let $E_1, E_2, \dots \subset \mathbb{R}^d$ be a countable sequence of sets. We want to show that

$$\inf \left\{ \sum_{n=1}^{\infty} |B_n| \mid B_n \text{ are boxes } \bigcup_{i=1}^{n} E_i \subset \bigcup_{n=1}^{\infty} Bn \right\} \leq \sum_{i=1}^{n} \inf \left\{ \sum_{n=1}^{\infty} \left| C_n^{(i)} \right| \mid C_n^{(i)} \text{ are boxes }, E_i \subset \bigcup_{n=1}^{\infty} C_n^{(i)} \right\} + \epsilon$$

for every $\epsilon > 0$. By the definition of infimum, for $\epsilon > 0$ we can find a cover $C_1^{(i)}, \dots$ of E_i such that $\sum_{n=1}^{\infty} \left| C_n^{(i)} \right| - m^*(E_i) \leq \epsilon / e^i$. Thus, by countable axiom of choice we can construct a double infinite sequence

$$\left(C_{(i,n)} \right)_{i \in \mathbb{N}, n \in \mathbb{N}} \text{ such that } E_i \subseteq \bigcup_{n=1}^{\infty} C_{(i,n)}$$

and

$$\sum_{n=1}^{\infty} \left| C_{(i,n)} \right| \leq m^*(E_i) + \epsilon/2^i.$$

But notice that

$$\bigcup_{i \in \mathbb{N}} E_i \subseteq \bigcup_{i \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} C_{(i,n)} = \bigcup_{(i,n) \in \mathbb{N}^2} C_{(i,n)} \implies \sum_{(i,n) \in \mathbb{N}^2} \left| C_{(i,n)} \right| \in \left\{ \sum_{n=1}^{\infty} |B_n| \mid B_n \text{ are boxes } \bigcup_{i=1}^{n} E_i \subset \bigcup_{n=1}^{\infty} B_n \right\}.$$

This satisfies

$$m^* \left(\bigcup_{i=1}^{n} E_i \right) \leq \sum_{(in) \in \mathbb{N}^2} \left| C_{(i,n)} \right| = \sum_{i \in \mathbb{N}} \sum_{n \in \mathbb{N}} \left| C_{(i,n)} \right| \leq \sum_{i \in \mathbb{N}} \left(m^*(E_i) + \epsilon/2^i \right) = \sum_{i \in \mathbb{N}} m^*(E_i) + \epsilon$$

as desired. □

Example. Let $E = \{x_1, \ldots\} \subset \mathbb{R}$ be a countable set. The Jordan outer measure of such a set can be arbitrarily large. For instance, the Jordan outer measure of $\mathbb{Q}$ is infinite. However, $\mathbb{Q}$ has Lebesgue outer measure 0, since we can cover it by the boxes $\{x_1\}, \{x_2\}, \ldots$, which each have volume 0. This shows that the Lebesgue outer measure of any countable set is 0.

To continue the analogy with the Jordan measure we would like to define a Lebesgue inner measure. However no information is gained from such a definition. So we will define the Lebesgue measure slightly differently.

Definition 22 (Lebesgue Measurable Sets)**.** A set $E \subset \mathbb{R}^d$ is said to be *Lebesgue measurable* if, for every $\epsilon > 0$, there exists an open set $U \subset \mathbb{R}^d$ containing E such that $m^*(U \backslash E) \leq \epsilon$. If E is Lebesgue measurable, we refer to $m(E) = m^*(E)$ as the *Lebesgue measure* of E.

Again we state the following without proof:

Theorem 23 ([Tao11] Lemma 1.2.13)**.** *The Lebesgue measure, m, satisfies the following:*

(a) (Empty set) $m(\emptyset) = 0$.

(b) (Monotonicity) If $E \subset F \subset \mathbb{R}^d$, then $m(E) \leq m(F)$.

(c) (Countable subadditivity) If $E_1, E_2, \ldots \subset \mathbb{R}^d$ is a countable sequence of sets, then

$$m \left(\bigcup_{n=1}^{\infty} E_n \right) \leq \sum_{n=1}^{\infty} m(E_n)$$

, and there is equality in the case where the E_i are pairwise disjoint.

Example. From the definitions we can see that every open set if Lebesgue measurable. So for $\mathbb{R}$ any interval of the form (a, b), or any arbitrary union of open intervals will be Lebesgue measurable. Also, the empty set is Lebesgue measurable, since we can just take arbitrarily small open sets.

Another example of a Lebesgue measurable set is a closed interval $[a, b]$. Given $\epsilon > 0$, there exists an $n \in \mathbb{N}$ such that $\frac{1}{n} < \epsilon$. Then we can take the set $E_n = (a - \frac{1}{2n}, b + \frac{1}{2n})$, and we will have

$$([a, b]) \backslash (a - \frac{1}{2n}, b + \frac{1}{2n}) = (a - \frac{1}{2n}, a) \cup (b, b + \frac{1}{2n})$$

which has Lebesgue outer measure $\frac{1}{2n}$. Thus since ϵ was arbitrary, $[a, b]$ is Lebesgue measurable.

So we define a set to be Lebesgue measurable if we can approximate it by open sets, in such a way that the error has very small Lebesgue outer measure.

We prove a quick lemma so that we can show that many sets are Lebesgue measurable.

Lemma 24 ([Tao11] Lemma 1.2.13)**.** *If $E \subseteq \mathbb{R}^d$ and $F \subseteq \mathbb{R}^d$ are Lebesgue measurable, then $E \backslash F$ is Lebesgue measurable.*

Proof. Given $\epsilon > 0$, there exists an open set $U \subset \mathbb{R}^d$ such that $m^*(U \backslash E) < \epsilon$. Then $U \backslash F$ contains $E \backslash F$ and is open, and this will approximate $E \backslash F$ with an arbitrarily small error, thus it is Lebesgue measurable. □

Lemma 25 ([Tao11] Lemma 1.2.13)**.** *If $E, F \subseteq \mathbb{R}^d$ are Lebesgue measurable sets, then $E_1 \cap E_2$ is Lebesgue measurable.*

Proof. For a given $\epsilon > 0$, there exist open sets $U, V \subseteq \mathbb{R}^d$ such that $m^*(U \backslash E) < \epsilon$ and $m^*(V \setminus F) < \epsilon$. Then $U_1 \cap U_2$ is open and will contain $E \cap F$ and this will approximate $E \cap F$ up to an arbitrary error. Thus $E \cap F$ is measurable. □

Proposition 26 ([Tao11] Lemma 1.2.13)**.** *If $E_1, E_2, \ldots, E_3 \subset \mathbb{R}^d$ are a sequence of Lebesgue measurable sets, then the union $\bigcup_{n=1}^{\infty} E_n$ is Lebesgue measurable.*

Proof. Let $\epsilon > 0$ be arbitrary. By assumption, each E_n is contained in an open set U_n whose difference $U_n \backslash E_n$ has Lebesgue outer measure at most $\epsilon/2^n$. By countable subadditivity, this implies that $\bigcup_{n=1}^{\infty} E_n$ is contained in $\bigcup_{n=1}^{\infty} U_n$ and the difference $(\bigcup_{n=1}^{\infty} U_n)\,(\bigcup_{n=1}^{\infty} E_n)$ has Lebesgue outer measure at most ϵ. The set $\bigcup_{n=1}^{\infty} U_n$ being a union of open sets, is itself open, and the claim follows. □

We end this section on the Lebesgue measure by showing that there are still sets that are not Lebesgue measurable.

Proposition 27 ([Tao11] Proposition 1.2.8)**.** *There exists a subset $E \subset [0, 1]$ which is not Lebesgue measurable.*

Proof. Consider the quotient $\mathbb{R}/\mathbb{Q}$. Each element of it is dense in $\mathbb{R}$ and thus has a non empty intersection with $[0, 1]$. Applying the axion of choice we can find an element $x_C \in C \cap [0, 1]$ for each $C \in \mathbb{R}/\mathbb{Q}$. We then let $E = \{x_C \mid C \in \mathbb{R}/\mathbb{Q}\}$ be the collection of coset representatives. By construction $E \subset [0, 1]$. Let y be any element of $[0, 1]$. Then it must lie in some coset C of $\mathbb{R}/\mathbb{Q}$, and thus differs from x_C by some rational number in $[-1, 1]$. In other words we have

$$[0, 1] \subset \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (E + q).$$

On the other hand, we have

$$\bigcup_{q \in \mathbb{Q} \cap [-1,1]} (E + q) \subset [-1, 2].$$

Also the different translates $E + q$ are disjoint, because E contains only one element from each coset of $\mathbb{Q}$.

We claim that E is not Lebesgue measurable. To see this suppose for the sake of contradiction that E was Lebesgue measurable. Then the translates $E + q$ would also be Lebesgue measurable. By countable additivity, we thus have

$$m\left(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} (E + q)\right) = \sum_{q \in \mathbb{Q} \cap [-1,1]} m(E + q),$$

and thus by translation invariance,

$$1 \leq \sum_{q \in \mathbb{Q} \cap [-11]} m(E) \leq 3.$$

On the other hand, the sum $\sum_{q \in \mathbb{Q} \cap [-1,1]} m(E)$ is either 0, if $m(E) = 0$, or infinite, if $m(E) > 0$, leading to a contradiction either way. □

Chapter 4

Lebesgue Integration

We define Lebesgue integration in a few steps. Starting with simple functions which play an analogous role to the piecewise constant functions we used in our discussion of Riemann integration.

Definition 28 (Characteristic Functions)**.** Let $A \subseteq \mathbb{R}^d$ be a measurable set. The *characteristic function* on A is defined as

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

Example. These are some of the simplest functions. Dirichlet's function can be defined as $\chi_{\mathbb{Q}\cap[0,1]}$.

Definition 29 (Simple Function)**.** A *simple function* $f\colon \mathbb{R}^d \to \mathbb{C}$ is a finite linear combination

$$f = c_1 1_{E_1} + \cdots + c_k 1_{E_k}$$

of indicator functions 1_{E_i} of Lebesgue measurable sets $E_i \subset \mathbb{R}^d$, and the $c_i \in \mathbb{C}$. An *unsigned simple function* $f\colon \mathbb{R}^d \to [0, +\infty]$ is defined similarly, but with the $c_i \in [0, +\infty]$ instead.

Example. Dirichlet's function that we discussed above is a simple function. Let $E_1 = \mathbb{Q} \cap [0, 1]$ and let $E_2 = \mathbb{R}\backslash\mathbb{Q} \cap [0, 1]$. Then Dirichlet's function is

$$f = 1 \cdot 1_{E_1} + 0 \cdot 1_{E_2}.$$

Any piecewise constant function is a simple function as well. Let $I = [0, 2]$ and consider the partition $\mathbb{P} = \{[0, 1], (1, 2], (2, 3]\}$. Then the function $f\colon I \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 2 & \text{if } x \in [0, 1] \\ 7 & \text{if } x \in (1, 2] \\ 1 & \text{if } x \in (2, 3] \end{cases}$$

is a simple function. We can write it as

$$f = 2 \cdot 1_{[0,1]} + 7 \cdot 1_{(1,2]} + 1 \cdot 1_{(2,3]}.$$

From this definition it is fairly intuitive for how we should define the Lebesgue integral of a simple unsigned function.

Definition 30 (Integral of Simple Functions)**.** If $f = c_1 1_{E_1} + \cdots + c_k 1_{E_k}$ is an unsigned simple function, the integral $\text{Simp}\int_{\mathbb{R}^d} f(x)dx$ is defined as

$$\text{Simp}\int_{\mathbb{R}^d} f(x)dx = c_1 m(E_1) + \cdots + c_k m(E_k).$$

Example. We can now take the unsigned simple integral of Dirichlet's function. Write it as $f(x) = 1 \cdot 1_{\mathbb{Q}\cap 1} + 0 \cdot 1_{\mathbb{R}\backslash\mathbb{Q}\cap[0,1]}$. Then we have

$$\text{Simp}\int_{[0,1]} f(x)dx = 1 \cdot m(\mathbb{Q} \cap [0,1]) + 0 \cdot m(\mathbb{R}\backslash\mathbb{Q} \cap [0,1]) = 1 \cdot m(\mathbb{Q} \cap [0,1]) = 0.$$

Definition 31 (Absolutely Integrable)**.** A complex simple function $f \colon \mathbb{R}^d \to \mathbb{C}$ is said to be *absolutely integrable* if $\text{Simp}\int_{\mathbb{R}^d} |f(x)|\,dx < \infty$. If f is absolutely integrable, the integral $\text{Simp}_{\mathbb{R}^d} f(x)dx$ is defined for real signed f by the formula

$$\text{Simp}\int_{\mathbb{R}^d} f(x)dx = \text{Simp}\int_{\mathbb{R}^d} f_+ dx - \text{Simp}\int_{\mathbb{R}^d} f_- dx$$

where $f_+ = \max(f(x), 0)$ and $f_-(x) = \max(-f(x), 0)$. And we define it for complex valued f by the formula

$$\text{Simp}\int_{\mathbb{R}^d} f(x)dx = \text{Simp}\int_{\mathbb{R}^d} \text{Re}\, f(x)dx + i\,\text{Simp}_{\mathbb{R}^d}\, \text{Im}\, f(x)dx$$

Definition 32 (Almost Everywhere)**.** We say that a property is said to hold *almost everywhere* if the set for which the property fails has Lebesgue measure zero.

Remark 33. The importance of almost every equality cannot be overstated. One of the crucial ideas of measure theory is that it does not matter if two functions, sets, etc. differ by a set of measure zero. We can consider them to be the same and everything works out nicely. This will come up over and over, some theorems will only hold for almost every $x \in \mathbb{R}^n$, we will identify to functions to be in the same equivalence class in some spaces if they are the same almost everywhere, and in many more places.

Example. We say that two functions, f and g are the same almost everywhere if the set $\{x \in \mathbb{R}^d \mid f(x) \neq g(x)\}$ has Lebesgue measure zero. So consider Dirichlet's function $f \colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}$$

Then the function $g(x) = 0$ is equal to $f(x)$ almost everywhere since the set where they are not equal is $\mathbb{Q} \cap [0,1]$ which has Lebesgue measure 0.

To define the Lebesgue integral of a general function we first need the notion of a measurable function, of which there are many, but we just choose one.

Definition 34 (Measurable Function)**.** An unsigned function $f \colon \mathbb{R}^d \to [0, +\infty]$ is *measurable* if for every open set $U \subset [0, +\infty]$, the set $f^{-1}(U)$ is Lebesgue measurable. An almost everywhere defined complex-valued function $f \colon \mathbb{R}^d \to \mathbb{C}$ is *measurable* if $f^{-1}(U)$ is measurable for every open set $U \subseteq \mathbb{C}$.

Example. If f is a continuous function, then $f^{-1}(U)$ is always open if U is open, so since every open set is Lebesgue measurable, f is measurable.

Consider Dirichlet's function $f\colon [0,1] \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

Let $U \subseteq [0,+\infty]$ be an open set, if $0 \notin U$ and $1 \notin U$, then $f^{-1}(U) = \emptyset$, which is measurable. If $0 \in U$ and $1 \in U$, then $f^{-1}(U) = [0,1]$, which is measurable since we can approximate it by the open sets $E_n = (-\frac{1}{n}, 1+\frac{1}{n})$ with arbitrarily small error. If $1 \in U$ and $0 \notin U$, then $f^{-1}(U) = \mathbb{Q} \cap [0,1]$, we have seen that $\mathbb{Q}$ is Lebesgue measurable and since $[0,1]$ is measurable, so is their intersection by lemma 25. Finally, if $0 \in U$ and $1 \notin U$, then $f^{-1}(U) = [0,1] \cap (\mathbb{R}\backslash\mathbb{Q})$, $\mathbb{R}$ is measurable since it is the complement of $\emptyset$, and since complements and intersections of measurable sets are measurable by lemmas 24 and 25, $f^{-1}(U)$ is measurable for all open $U \subseteq [0,+\infty]$ and thus f is measurable.

Definition 35 (Lower and Upper Lebesgue Integral)**.** Let $f\colon \mathbb{R}^d \to [0,+\infty]$ be an unsigned function. We define the *lower unsigned Lebesgue integral* $\underline{\int_{\mathbb{R}^d}} f(x)dx$ to be the quantity

$$\underline{\int_{\mathbb{R}^d}} f(x)dx = \sup\left\{\mathrm{Simp}\int_{\mathbb{R}^d} g(x)dx \mid 0 \le g \le f; g \text{ simple}\right\}$$

where g ranges over all unsigned simple functions $g\colon \mathbb{R}^d \to [0,+\infty]$ that are pointwise bounded by f. There is an analogous definition of upper unsigned Lebesgue integral but we will not use it.

Now we have

Definition 36 (Unsigned Lebesgue Integral)**.** If $f\colon \mathbb{R}^d \to [0,+\infty]$ is measurable, we define the *unsigned Lebesgue integral* $\int_{\mathbb{R}^d} f(x)dx$ of f to be equal to the lower unsigned Lebesgue integral.

Finally we get to the notion of absolute integrability.

Definition 37 (Absolutely Integrable)**.** An almost everywhere defined measurable function $f\colon \mathbb{R}^d \to \mathbb{C}$ is said to be *absolutely integrable* if the unsigned integral

$$\int_{\mathbb{R}^d} |f(x)|\,dx$$

is finite. If f is real-valued and absolutely integrable, we define te Lebesgue integral by the formula

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} f_+(x)dx - \int_{\mathbb{R}^d} f_-(x)dx,$$

and if f is complex valued we define it by

$$\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} \mathrm{Re}\, f(x)dx + i\int_{\mathbb{R}^d} \mathrm{Im}\, f(x)dx.$$

Now we return to the function $f\colon [0,1] \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is rational} \\ 0 & \text{if } x \text{ is irrational} \end{cases}.$$

We saw that this function is not Riemann integrable. However, this is just the characteristic function on $\mathbb{Q} \cap [0,1]$, which is a measurable set, so it is Lebesgue integrable, with Lebesgue integral 0, since this set has Lebesgue measure 0.

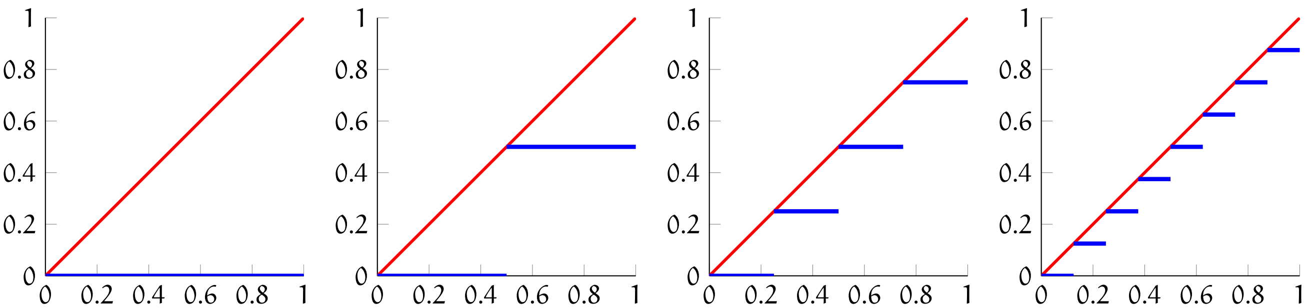


Figure 4.1: Approximation of $f(x) = x$ by simple functions from below.

Now we prove that if a function is Riemann integrable, then it is Lebesgue integrable and the integrals are equal. To distinguish between the two integrals we denote the Riemann integral by $R\int_I f$.

To prove our last proposition we make a quick definition that we could have made before, but it will only be used here.

Definition 38 (Upper and Lower Lebesgue Integral)**.** Let E be a measurable subset of $\mathbb{R}^n$, and let $f\colon E \to \mathbb{R}$ be a function. We define the *upper Lebesgue integral* to be

$$\overline{\int_E} f = \inf\left\{\int_E g \mid g \text{ is an absolutely integrable function from } E \text{ to } \mathbb{R} \text{ that majorizes } f\right\}$$

and we define the *lower Lebesgue integral* to be

$$\underline{\int_E} f = \sup\left\{\int_E g \mid g \text{ is an absolutely integrable function from } E \text{ to } \mathbb{R} \text{ that minorizes } f\right\}$$

Example. Consider the function $f(x) = x^3$ on the interval $[0, 1]$. We see that the function $g(x) = x$ majorizes f since $f(x) = x^3 \leq x$ on $[0, 1]$. Furthermore, we have

$$\int_{[0,1]} |x|\, dx = \int_{[0,1]} x dx = \sup\left\{\int_{[0,1]} s \mid s \text{ is simple and non-negative and minorizes } x\right\}.$$

We can see that simple functions approximate the function $g(x) = x$. Notice that in this case that since x is Riemann integrable, we can approximate it by piecewise constant functions. So $\int_{[0,1]} x dx = \frac{1}{2}$ and this function majorizes $f(x) = x^3$, so we get

$$\int_{[0,1]} x^3 dx \leq \int_{[0,1]} x dx.$$

Taking absolutely integrable functions closer and closer to $f(x)$ will get us an upper bound on it. Similarly, we can approximate it from below by the lower Lebesgue integral and get a lower bound on the function. In this case, as we will since in the next proposition, since the function is Riemann integrable, it is Lebesgue integral and its Riemann and Lebesgue integrals are equal.

Then we have

Lemma 39 ([Tao15b] Lemma 8.3.6)**.** *Let* E *be a measurable subset of* $\mathbb{R}^d$ *and let* $f\colon E \to \mathbb{R}$ *be a function. Let* A *be a real number, and suppose* $\overline{\int}_E f = \underline{\int}_E f = A$*. Then* f *is absolutely integrable and*

$$\int_E f = \overline{\int_E} f = \underline{\int_E} f = A.$$

Finally, we have:

Proposition 40 ([Tao15b] Proposition 8.4.1)**.** *Let* $I \subseteq \mathbb{R}$ *be an interval, and let* $f\colon I \to \mathbb{R}$ *be a Riemann integrable function. Then* f *is also absolutely integrable, and* $\int_I f = R\int_I f$*.*

Proof. Write $A = R\int_I f$. Since f is Riemann integrable, we know that the upper and lower Riemann integrals are equal to A. Thus, for every $\epsilon > 0$, there exists a partition $\mathbb{P}$ of I into smaller intervals J such that

$$A - \epsilon \leq \sum_{J \in \mathbb{P}} |J| \inf_{x \in J} f(x) \leq A \leq \sum_{J \in \mathbb{P}} |J| \sup\{f(x) x \in J\} \leq A + \epsilon.$$

Let $f_\epsilon^-\colon I \to \mathbb{R}$ be the function

$$f_\epsilon^- = \sum_{J \in \mathbb{P}} \inf\{f(x) \mid x \in J\} 1_J(x)$$

and let f_ϵ^+ be defined as

$$f_\epsilon^+ = \sum_{J \in \mathbb{P}} \sup\{f(x) \mid x \in J\} 1_J(x).$$

These are simple function and hence are measurable and absolutely integrable. We have

$$\int_I f_\epsilon^- = \sum_{J \in \mathbb{P}} \inf\{f(x) \mid x \in J\} 1_J(x)$$

and

$$\int_I f_\epsilon^+ = \sum_{J \in \mathbb{P}} \sup\{f(x) \mid x \in J\} 1_J(x).$$

And hence

$$A - \epsilon \leq \int_I f_\epsilon^- \leq A \leq \int_I f_\epsilon^+ \leq A + \epsilon.$$

Since f_ϵ^+ majorizes f and f_ϵ^- minorizes f, we thus have

$$A - \epsilon \leq \underline{\int_I} f \leq \overline{\int_I} f \leq A + \epsilon$$

for every ϵ, and thus

$$\underline{\int_I} f = \overline{\int_I} f = A$$

and thus f is absolutely integrable $\int_I f = A$. □

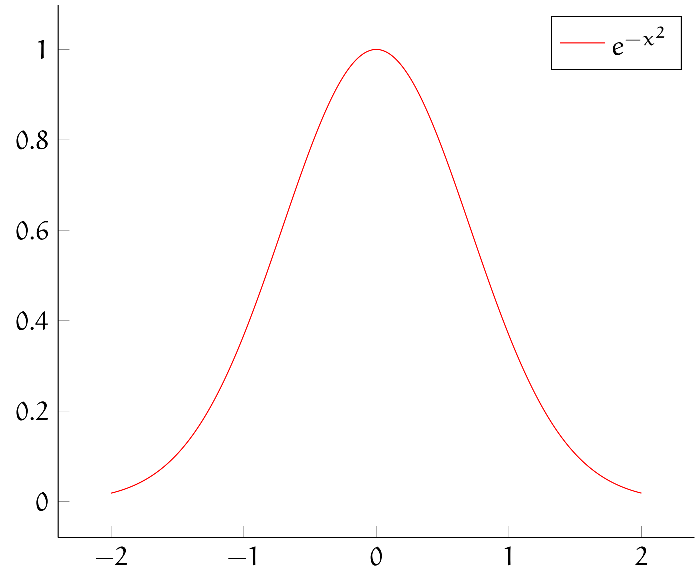


Figure 4.2: The graph of e^{-x^2}

We end with a brief geometric discussion on the difference between the Riemann integral and the Lebesgue integral. The Riemann integral can be thought of as taking the area under the curve by approximating it with boxes, whose width goes to 0. This works well for many functions, but can easily fail. The difference with the Lebesgue integral is it switches the perspective to use horizontal boxes. For each value that the function takes it is basically multiplying that value by the size of the set that takes the value, which is its measure. The idea of measure of a set was what allowed Lebesgue to define the integral in this different, and better way. The following example illustrates this idea.

Example. If we wanted to evaluate the Riemann integral of this function, then we would begin by partitioning the domain, say we partition the interval $[-2,2]$ as $\mathbb{P} = \{[-2,-1],(-1,0],(0,1],(1,2]\}$. Then we can take the Upper Riemann sum with respect to this partition to be

$$U(e^{-x^2},\mathbb{P}) = \sum_{J\in\mathbb{P}} (\sup\{f(x) \mid x \in J\}) \cdot |J| .$$

Then to evaluate the Riemann integral we will shrink down the length of the intervals in the partition to 0. On the other hand, if we want to evaluate the Lebesgue integral, we will start by partitioning the range of the function. We can pick values on the range such that $\{0.2, 0.4, 0.6, 0.8, 1\}$, and then create the simple function with respect to these values. Then for 0.2 we would take the function $0.2 \cdot 1_{A_{0.2}}$ where $A_{0.2}$ is the subset of the interval [-2, 2] where

$$A_{0.2} = [-2,x] \cup [x',2]$$

where x is the smallest point such that $f(x) = 0.2$ and x' is the largest point such that $f(x') = 0.2$. Then we would make $A_{0.4}$ which is defined as

$$A_{0.4} = [x,y] \cup [y',x']$$

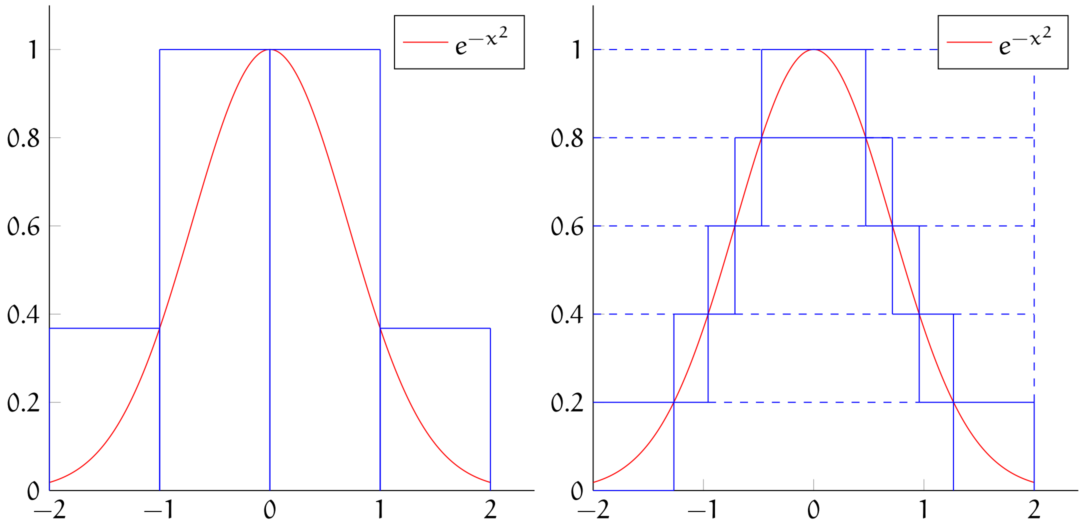


Figure 4.3: The Riemann and Lebesgue integral of e^{-x^2}, respectively.

where x and x' are as above and y is the smallest point such that $f(y) = 0.4$ and y' is the largest point such that $f(y') = 0.4$. We can continue in this way to get the rest of the sets $A_{0.6}, A_{0.8}, A_1$. Then our simple function is

$$0.2 \cdot 1_{A_{0.2}} + 0.4 \cdot 1_{A_{0.4}} + 0.6 \cdot 1_{A_{0.6}} + 0.8 \cdot 1_{A_{0.8}} + 1 \cdot 1_{A_1}.$$

We can see this in second image of figure 4.3. On the other hand, if we wanted to integrate this function with the Riemann integral, we would begin by partitioning the domain. We might partition $[-2, 2]$ as

$$\mathbb{P}' = \{[-2,-1], (-1,0], (0,1], (1,2]\}.$$

Then we can take the piecewise function that majorizes e^{-x^2}, by taking the value on each piece of the partition to be the supremum of values on the partition, as can be seen in the left hand side of figure 4.3.

The main difference between the two methods of integration is that with the Lebesgue integral we are able to break up the domain into more pieces, and pieces that are not connected. Whereas with the Riemann integral, we must partition the domain first, but as we have seen this does not work in the case of more complex functions such as Dirichlet's function. In that case the partition of the domain is complex, while the range is particularly simple, it only assumes two values on the range. For functions such as these the Lebesgue integral works far better.

4.1 Convergence Theorems

In this section we prove the monotone convergence theorem, Fatou's lemma, and the dominated convergence theorem because all three are frequently used throughout this thesis. These theorems give conditions for when a limit and an integral commute. This is another advantage of Lebesgue integration over Riemann integration. This section follows [Bas13]. Throughout we let μ denote the Lebesgue measure.

Theorem 41 (Monotone Convergence Theorem, [Bas13] Theorem 7.1)**.** *Suppose f_n is a sequence of non-negative measurable functions with $f_1(x) \leq f_2(x) \leq \cdots$ for all x and with*

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

Proof. Note that $\int f_n$ is an increasing sequence. Let L be the limit of this sequence. Since $f_n \leq f$ for all n, we have $L \leq \int f$, we show the opposite inequality.

Let $s = \sum_{i=1}^{m} a_i \chi_{E_i}$ be any non-negative simple function less than or equal to f and let $c \in (0,1)$. Let $A_n = \{x \mid f_n(x) \geq cs(x)\}$. Since $f_n(x)$ increases to $f(x)$ for each x and $c < 1$, then $A_1 \subseteq A_2 \subseteq \cdots$ and $\bigcup_{n=1}^{\infty} A_n = \mathbb{R}^n$. For each n,

$$\begin{aligned} \int f_n \geq \int_{A_n} f_n \geq c \int_{A_n} s \\ = c \int_{A_n} \sum_{i=1}^{m} a_i \chi_{E_i} \\ = c \sum_{i=1}^{m} a_i \mu(E_i \cap A_n). \end{aligned}$$

If we let $n \to \infty$ then the right hand side converges to

$$c \sum_{i=1}^{m} a_i \mu(E_i) = c \int s.$$

Therefore, $L \geq c \int s$. Since c is arbitrary in the interval $(0,1)$, then $L \geq \int s$. Take the supremum over all simple $s \leq f$, we obtain $L \geq \int f$. □

Example.

(a) Let $f\colon \mathbb{R}^n \to [0,\infty]$ be a non-negative function. We will prove that $\lim_{n\to\infty} \int_{\mathbb{R}^n} n \log\left(1 + \frac{f}{n}\right) d\mu = \int_{\mathbb{R}^n} f d\mu$. First, let

$$f_n = n \log\left(1 + \frac{f}{n}\right) = \log\left(1 + \frac{f}{n}\right)^n.$$

Note that each n is non-negative and measurable. Further, $f_1 \leq f_2 \leq \cdots$ since $\log$ is an increasing function and the sequence

$$\lim_{n\to\infty}\left(1 + \frac{f(x)}{n}\right)^n = e^{f(x)}$$

is an increasing sequence. Thus by the monotone convergence theorem

$$\lim_{n\to\infty} \int_{\mathbb{R}^n} f_n d\mu = \int_{\mathbb{R}^n} f d\mu.$$

(b) Let our space be $X = [0,\infty)$ with the Lebesgue measure and take $f_n(x) = -1/n$. Then $\int_X f_n = -\infty$, but $f_1 \leq f_2 \leq \cdots \leq f = 0$ and $\int_X 0 = 0$. The monotone convergence theorem does not work here because the f_n are negative.

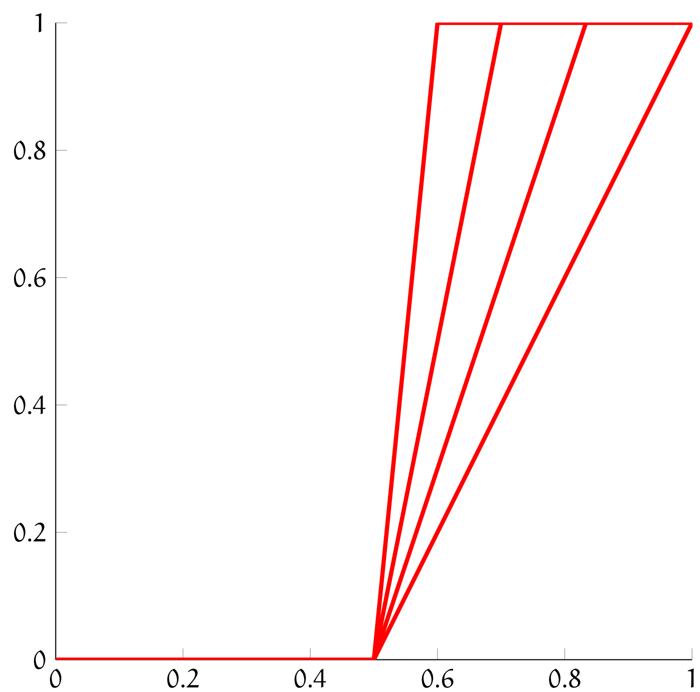


Figure 4.4: An increasing sequence of functions.

(c) Suppose $f_n = n\chi_{(0,1/n)}$ .Then $f_n \geq 0$, $f_n \to 0$ for each x, but $\int_X f_n = 1$ does not converge to $\int 0 = 0$. The monotone convergence theorem does not work here because the f_n decrease to f, they don't increase to f.

Remark 42. The monotone convergence theorem gives us a condition for when lim and $\int$ commute. The next theorem, Fatou's lemma tells us the best we can do when we don't put any conditions on the f_n.

Theorem 43 (Fatou's Lemma, [Bas13] Theorem 7.8)**.** *Suppose the f_n are non-negative and measurable. Then*

$$\int_{\mathbb{R}^n} \liminf_{n\to\infty} f_n \leq \liminf_{n\to\infty} \int_{\mathbb{R}^n} f_n.$$

Proof. Let $g_n = \inf_{i\geq n} f_i$. Then the g_n are non-negative and g_n increases to $\liminf_{n\to\infty} f_n$. Clearly $g_n \leq f_i$ for each $i \geq n$, so $\int g_n \leq \int f_i$. Therefore

$$\int g_n \leq \int_{i\geq n} \int f_i.$$

If we take the limit in the above inequality as $n \to \infty$ on the left hand side we obtain $\int \liminf f_n$ by the monotone convergence theorem and on the right hand side we obtain $\liminf \int f_n$. □

Example. Let $E \subset \mathbb{R}$ be Lebesgue measurable and define

$$f_n = \begin{cases} \chi_E & \text{if } n \text{ is even} \\ 1 - \chi_E & \text{if } n \text{ is odd.} \end{cases}$$

Consider what happens if $X = [0,2]$ and $E = (1,2]$. Then we get the sequence of functions

$$f_n = \begin{cases} \chi_{(1,]} & \text{if } n \text{ is even} \\ \chi_{[0,1]} & \text{if } n \text{ is odd.} \end{cases}$$

Note that as n increases, the graphs of these functions swap back and forth and that for any n

$$\int_{[0,2]} f_n = 1$$

but $\liminf_n f_n = 0$. Thus

$$0 = \int_{[0,2]} \liminf_{n\to\infty} f_n < \liminf_{n\to\infty} \int_{[0,2]} f_n = 1$$

which is an example of strict inequality in Fatou's lemma.

Finally we reach the dominated convergence theorem which is similar to the monotone convergence theorem in that it gives us a condition for which lim and $\int$ commute.

Theorem 44 (Dominated Convergence Theorem, [Bas13] Theorem 7.9)**.** *Suppose that f_n are measurable real-valued functions and $f_n(x) \to f(x0$ for each x. Suppose that there exists a non-negative integrable function g such that $|f_n(x)| \leq g(x)$ for all x. Then*

$$\lim_{n\to\infty} \int_{\mathbb{R}^n} f_n \, d\mu \to \int f \, d\mu.$$

Proof. Since $f_n + g \geq 0$, by Fatou's lemma,

$$\int f + \int g = \int (f+g) \leq \liminf_{n\to\infty} \int (f_n + g) = \liminf_{n\to\infty} \int f_n + \int g.$$

Since g is integrable,

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

Similarly, $g - f_n \geq 0$, so

$$\int g - \int f = \int (g - f) \leq \liminf_{n\to\infty} \int (g - f_n) = \int f + \liminf_{n\to\infty} \int (-f_n),$$

and hence

$$-\int f \leq \liminf_{n\to\infty} \int (-f_n) = -\limsup_{n\to\infty} f_n.$$

Therefore

$$\int f \geq \limsup_{n\to\infty} \int f_n$$

which proves the theorem. □

Example. We will use the Dominated Convergence theorem to compute

$$\lim_{n\to\infty} \int_{\mathbb{R}} \frac{n \sin(x/n)}{x(x^2+1)} \, d\mu.$$

Let $x \in \mathbb{R}$ and define

$$f_n(x) = \frac{n\sin(x/n)}{x(x^2+1)}$$

for each $n \in \mathbb{N}$. Note that each f_n is measurable and the sequence $\{f_n\}$ converges pointwise to $\frac{1}{1+x^2}$ for all $x \neq 0$. Indeed,

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty}\left(\frac{\sin(x/n)}{x/n}\right)\frac{1}{1+x^2} = \frac{1}{1+x^2}.$$

This allows us to choose $g(x) = \frac{1}{1+x^2}$ as our dominating function. Thus we can apply the dominated convergence theorem to conclude that

$$\begin{aligned}
\lim_{n\to\infty}\int_{\mathbb{R}} \frac{n\sin(x/n)}{x(x^2+1)}\,d\mu &= \lim_{n\to\infty}\int_{-\infty}^{\infty} \frac{n\sin(x/n)}{x(x^2+1)}\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx \\
&= \arctan(x)|_{-\infty}^{\infty} \\
&= \pi.
\end{aligned}$$

Chapter 5

Abstract Measure Theory

In this section we very briefly introduce abstract measure theory with the goal of proving the Radon-Nikodym Theorem. This theorem is crucial in the proof of the duality of L^p and L^q. Everything in this section is based on [Bas13] but can also be found in [Tao11], [Roy88] or any other standard analysis book. Most of the definitions and theorems here are exactly the same as those in the Lebesgue measure section, just in a more general setting. Throughout this section let X be a set. Taking $X = \mathbb{R}^n$ we will be in the Lebesgue setting.

Definition 45. A σ-algebra is a collection $\mathcal{A}$ of subsets of X such that

(a) $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$.

(b) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.

(c) Whenever $A_1, A_2, \ldots$ are in $\mathcal{A}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$ and $\bigcap_{i=1}^{\infty} A_i \in \mathcal{A}$.

An example of this we have already seen is the collection of Lebesgue measurable sets.

Definition 46. The pair $(X, \mathcal{A})$ is called a *measurable space* and a set A is *measurable* if $A \in \mathcal{A}$.

Remark 47. A typical example of a measurable space is a probability space. One can think of a σ-algebra as the possible events in the space, or as the set of possible outcomes of an experiment.

Definition 48. Let $(X, \mathcal{A})$ be a measurable space. A *measure* on $(X, \mathcal{A})$ is a function $\mu \colon \mathcal{A} \to [0, \infty]$ such that

(a) $\mu(\emptyset) = 0$.

(b) If $A_i \in \mathcal{A}$, $i = 1, 2, \ldots$ are pairwise disjoint, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

We call the triple $(X, \mathcal{A}, \mu)$ a *measure space*.

These are the same properties that we wanted the Lebesgue measure to satisfy. We are simply taking those ideas and abstracting them.

Example. Let $X = \mathbb{N}$, let $\mathcal{A} = 2^X$, the power set of X and let $\mu(A) = |A|$ for $A \subseteq X$. This is called the *counting measure* of X and simply counts how many elements are in the given set.

Definition 49. Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that μ is a *finite measure* if $\mu(X) < \infty$. We also say that μ is σ-finite if there exist sets $E_i \in \mathcal{A}$ such that $\mu(E_i) < \infty$ for each i and $X = \bigcup_{i=1}^{\infty} A_i$.

Example. Consider the measure space $\mathbb{R}$ with the Lebesgue measure. Let $E_i = (-i, i)$. Then for each i, $\mu(E_i) < infty$, but $\bigcup_{i=1}^{\infty} E_i = \mathbb{R}$, so the Lebesgue measure is a σ-finite measure.

Next we define the concept of measurable function which is the same as in the case of the Lebesgue measure. This will allow us to define the integral of a measurable function with respect to a particular measure.

Definition 50. A function $f\colon X \to \mathbb{R}$ is *measurable* if $\{x \mid f(x) > a\} \in \mathcal{A}$ for all $a \in \mathbb{R}$. A complex-valued function is measurable if both its real and imaginary parts are measurable.

Example. Suppose f is real valued and constant. Then the set $\{x \mid f(x) > a\}$ is either empty or all of X, so f is measurable.

Definition 51. Let $(A, \mathcal{A}, \mu)$ be a measure space. If

$$s = \sum_{i=1}^{n} a_i \chi_{E_i}$$

is a non-negative measurable simple function, the integral of s is

$$\int_X s \, d\mu = \sum_{i=1}^{n} a_i \mu(E_i).$$

Here, if $a_i = 0$ and $\mu(E_i) = \infty$, we use the convention that $a_i \mu(E_i) = 0$. If $f \geq 0$ is a measurable function, define

$$\int_X f \, d\mu = \sup \left\{ \int s \, d\mu \mid 0 \leq s \leq f, s \text{ simple} \right\}.$$

Finally, to define the integral of a general measurable function let f be measurable and set $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$. Provided that $\int_X f^+ d\mu$ and $\int_X f^- d\mu$ are both finite, define

$$\int_X f \, d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

If $f = u + iv$ is a complex valued function define

$$\int_X f \, d\mu = \int_X u \, d\mu + i \int_X v \, d\mu.$$

One of the most important propositions in this section is the following criteria for a function to be zero almost everywhere. This will be used constantly.

Proposition 52. *Suppose* f *is measurable and non-negative and* $\int f \, d\mu = 0$. *Then* $f = 0$ *almost everywhere.*

Proof. If f is not equal to 0 almost everywhere, there exists an n such that $\mu(A_n) > 0$ where $A_n = \{x \mid f(x) > 1/n\}$. But since f is non-negative,

$$0 = \int f \geq \int_{A_n} f \geq \frac{1}{n} \mu(A_n),$$

a contradiction. □

This will prove to be particularly useful when we are looking at the integrals of $|f|$.

5.1 Signed Measures

Now we proceed towards the Radon-Nikodym Theorem. We will first look at signed measures which are a measure that takes negative values as well as positive.

Definition 53 (Signed Measure)**.** Let $\mathcal{A}$ be a σ-algebra. A *signed measure* is a function $\mu\colon (-\infty, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever the A_i are pairwise disjoint.

Definition 54 (Positive, Negative, Null Sets)**.** Let μ be a signed measure. A set $A \subset \mathcal{A}$ is called a *positive set* for μ if $\mu(B) \geq 0$ whenever $B \subset A$ and $B \in \mathcal{A}$. We say that $A \in \mathcal{A}$ is a *negative* set if $\mu(B) \leq 0$ whenever $B \subset A$ and $B \in \mathcal{A}$. Finally, a *null set* is one where $\mu(B) = 0$ whenever $B \subset A$ and $B \in \mathcal{A}$.

Example. Suppose μ is the Lebesgue measure and

$$\nu(A) = \int_A f d\mu$$

for some integrable f. If we let $P = \{x \mid f(x) \geq 0\}$, then P is a positive set and if $N = \{x \mid f(x) \leq 0\}$, then N is a negative set. The Hahn decomposition theorem below will decompose our space into P and N. First we need a proposition.

Proposition 55 ([Bas13] Proposition 12.4)**.** *Let μ be a signed measure which takes values in $(-\infty, \infty]$. Let E be a measurable set with $\mu(E) < 0$. Then there exists a measurable subset F of E that is a negative set with $\mu(F) < 0$.*

Proof. If E is a negative set, we are done. If not, there exists a measurable subset with positive measure. Let n_1 be the smallest positive integer such that there exists $E_1 \subset E$ with $mu(E_1) \geq 1/n_1$. We then define pairwise disjoint measurable sets $E_2, E_3, \ldots$ by induction as follows. Let $k \geq 2$ and suppose $E_1, \ldots, E_{k-1}$ are pairwise disjoint measurable sets with $\mu(E_i) > 0$ for $i = 1, \ldots, k-1$. If $F_k = E - (E_1 \cup \cdots \cup E_{k-1}$ is a negative set, then

$$\mu(F_k) = \mu(E) - \sum_{i=1}^{k-1} \mu(E_i) \leq \mu(E) < 0$$

and F_k is the desired set F. If F_k is not a negative set, let n_k be the smallest positive integer such that there exists $E_k \subset F_k$ with E_k measurable and $\mu(E_k) \geq 1/n_k$. We stop the construction if there exists a k such that F_k is a negative set with $\mu(F_k) < 0$. If not, we continue and let $F = \bigcap_k F_k - E - (\bigcup_k E_k)$. Since $0 > \mu(E) > -\infty$ and $\mu(E_k) \geq 0$, then

$$\mu(E) = \mu(F) + \sum_{k=1}^{\infty} \mu(E_k).$$

Then $\mu(F) \leq \mu(E) < 0$, so the sum converges.

Lastly we show that F is a negative set. Suppose $G \subset F$ is measurable with $\mu(G) > 0$. Then $\mu(G) \geq 1/N$ for some N. But this contradicts the construction, since for some k, $n_k > N$, and we would have chosen the set G instead of the set E_k at stage k. Therefore F must be a negative set. □

Now we arrive at the Hahn Decomposition Theorem which will be a crucial part of the Radon-Nikodym theorem. We write $A \triangle B$ for $(A - B) \cup (B - A)$.

Theorem 56 (Hahn Decomposition Theorem, [Bas13] Theorem 12.5)**.**

(a) Let μ be a signed measure taking values in $(-\infty, \infty]$. There exist disjoint measurable sets E and F in $\mathcal{A}$ whose union is X and such that E is a negative set and F is a positive set.

(b) If E' and F' are another such pair, then $E \triangle E' = F \triangle F'$ is a null set with respect to μ.

(c) If μ is not a positive measure, then $\mu(E) < 0$. If $-\mu$ is not a positive measure, then $\mu(F) > 0$.

Proof.

(a) Let $L = \inf\{\mu(A) \mid A \text{ is a negative set}\}$. Choose negative sets A_n such that $\mu(A_n) \to L$. Let $E = \bigcup_{n=1}^{\infty} A_n$. Let $B_n = A_n - (B_1 \cup \cdots \cup B_{n-1})$ for each n. Since A_n is a negative set, so is each B_n. Also, the B_n are disjoint and $\bigcup_n B_n = \bigcup_n A_n = E$. If $C \subset E$, then

$$\mu(C) = \lim_{n\to\infty} \mu\left(C \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(C \cap B_i) \leq 0.$$

Thus E is a negative set. Since E is a negative set,

$$\mu(E) = \mu(A_n) + \mu(E - A_n) \leq \mu(A_n).$$

Letting $n \to \infty$, we obtain $\mu(E) = L$.

Let $F = E^c$. If F were not a positive set, there would exist $B \subset F$ with $\mu(B) < 0$. By Proposition 55 there exists a negative set C contained in B with $\mu(C) < 0$. But then $E \cup C$ would be a negative set with $\mu(E \cup C) < \mu(E) = L$, a contradiction.

(b) To prove uniqueness, if E', F' are another such pair of sets and $A \subset E - E' \subset E$, then $\mu(A) \leq 0$. But $A \subset A - E' = F' - F \subset F'$, so $\mu(A) \geq 0$. Therefore $\mu(A) = 0$. The same argument works if $A \subset E' - E$, and any subset of $E \triangle E'$ can be written as the union of A_1 and A_2 where $A_1 \subset E - E'$ and $A_2 \subset E' - E$.

(c) Suppose μ is not a positive measure but $\mu(E) = 0$. If $A \in \mathcal{A}$, then

$$\mu(A) = \mu(A \cap E) + \mu(A \cap F) \geq \mu(E) + \mu(A \cap F) \geq 0,$$

which says that μ must be a positive measure, a contradiction. A similar argument applies for $-\mu$ and F.

□

This theorem leads to the following definition.

Definition 57. We say that two measures are *mutually singular* if there exists two disjoint sets E and F in $\mathcal{A}$ such that $E \cup F = X$ and $\mu(E) = \mu(F) = 0$. This is denoted $\mu \perp \nu$.

Suppose f is non-negative and integrable with respect to μ. If we define ν by

$$\nu(A) = \int_A f \, d\mu,$$

then ν is a measure. We will now consider the converse. If we are given two measures ν and μ, when can we find an f such that $\nu(A) = \int_A f \, d\mu$ holds for all subsets $A \in \mathcal{A}$.

Definition 58. A measure ν is said to be *absolutely continuous* with respect to a measure μ if $\nu(A) = 0$ whenever $\mu(A) = 0$. We write $\nu \ll \mu$.

We need one last lemma before we can prove the Radon-Nikodym theorem.

Lemma 59 ([Bas13] Lemma 13.3)**.** *Let μ and ν be finite positive measures on a measurable space $(X, \mathcal{A})$. Either $\mu \perp \nu$ or else there exists $\epsilon > 0$ and $G \in \mathcal{A}$ such that $\mu(G) > 0$ and G is a positive set for $\nu - \epsilon\nu$.*

Proof. Consider the Hahn decomposition for $nu - \frac{1}{n}\mu$. Thus there exists a negative set E_n and a positive set F_n for this measure, E_n and F_n are disjoint and their union is X. Let $F = \bigcup_n F_n$ and $E_n = \bigcap_n E_n$. Note $E^c = \bigcup_n E_n^c = \bigcup_n F_n = F$.

For each $n, E \subset E_n$, so

$$\nu(E) \leq \nu(E_n) \leq \frac{1}{n}\mu(E_n) \leq \frac{1}{n}\mu(X).$$

Since ν is a positive measure this implies that $\nu(E) = 0$. One possibility is that $\mu(E^c) = 0$, in which case $\mu \perp \nu$. The other possibility is that $\mu(E^c) > 0$. In this case, $\mu(F_n) > 0$ for some n. Let $\epsilon = \frac{1}{n}$ and $G = F_n$. Then from the definition of F_n, G is a positive set for $\nu - \epsilon\mu$. □

Finally we can state and prove the Radon-Nikodym theorem.

Theorem 60 (Radon-Nikodym, [Bas13] Theorem 13.4)**.** *Suppose μ is a σ-finite positive measure on a measurable space $(X, \mathcal{A})$ and ν is a finite positive measure on $(X, \mathcal{A})$ such that ν is absolutely continuous with respect to μ. Then there exists a μ-integrable non-negative function f which is measurable with respect to $\mathcal{A}$ such that*

$$\nu(A) = \int_A f d\mu$$

for all $A \in \mathcal{A}$. Moreover, if g is another such function, then $f = g$ almost everywhere with respect to μ.

Proof. The main idea of the proof is to look at the set of f such that $\int_A f d\mu \leq \nu(A)$ and choose the f such that the integral is largest. However, we will start by proving the uniqueness statement. Suppose f and g are two functions such that

$$\int_A f d\mu = \nu(A) = \int_A g d\mu$$

for all $A \in \mathcal{A}$. For every set A we have

$$\int_A (f - g) = \nu(A) - \nu(A) = 0$$

but this implies that $f - g = 0$ almost everywhere with respect to μ.

Next let us assume that μ is a finite measure. Define

$$\mathcal{F} = \left\{ g \text{ measurable } \mid g \geq 0, \int_A g d\mu \leq \nu(A) \text{ for all } A \in \mathcal{A} \right\}.$$

Note that $\mathcal{F}$ is not empty because $0 \in \mathcal{F}$. Let $L = \sup\left\{\int g d\mu \mid g \in \mathcal{F}\right\}$, and let g_n be a sequence in $\mathcal{F}$ such that $\int g_n d\mu \to L$. Let $h_n = \max(g_1, \ldots, g_n)$.

We claim that if $g_1, g_2 \in \mathcal{F}$, then $h_2 = \max(g_1, g_2) \in \mathcal{F}$ as well. To see this, let $B = \{x \mid g_1(x) \geq g_2(x)\}$, and write

$$\begin{aligned}
\int_A h_2 d\mu &= \int_{A \cap B} h_2 d\mu + \int_{A \cap B^c} h_2 d\mu \\
&= \int_{A \cap B} g_1 d\mu + \int_{A \cap B^c} g_2 d\mu \\
&\leq \nu(A \cap B) + \nu(A \cap B^c) \\
&= \nu(A).
\end{aligned}$$

Therefore $h_1 \in \mathcal{F}$. By induction, one can also show that $h_n \in \mathcal{F}$.

The h_n increase to some function f. By the monotone convergence theorem,

$$\int_A f d\mu \leq \nu(A)$$

for all $A \in \mathcal{A}$ and

$$\int f d\mu \geq \int h_n d\mu \geq \int g_n d\mu$$

for each n, so $\int f d\mu = L$.

Next we prove that f is the desired function. Define a measure λ by

$$\lambda(A) = \nu(A) - \int_A f d\mu.$$

Note that λ is a positive measure since $f \in \mathcal{F}$. Suppose λ is not mutually singular to μ. By Lemma 59, there exits $\epsilon > 0$ and $G \in \mathcal{A}$, $\mu(G) > 0$ and G is a positive set for $\lambda - \epsilon\mu$. For any $A \in \mathcal{A}$,

$$\nu(A) - \int_A f d\mu = \lambda(A) \geq \lambda(A \cap G) \geq \epsilon\mu(A \cap G) = \int_A \epsilon G_{\chi_G} d\mu,$$

or

$$\nu(A) \geq \int_A (f + \epsilon\chi_G) d\mu.$$

Hence $f + \epsilon\chi_G \in \mathcal{F}$. But

$$\int_X (f + \epsilon\chi_G) d\mu = L + \epsilon\mu(G) > L,$$

a contradiction to the definition of L. Therefore $\lambda \perp \mu$. Then there must exist $H \in \mathcal{A}$ such that $\mu(H) = 0$ and $\lambda(H^c) = 0$. Since $\nu \ll \mu$, then $\nu(H) = 0$, and hence

$$\lambda(H) = \nu(H) - \int_H f d\mu = 0.$$

This implies that $\lambda(A) = 0$ or $\nu(A) = \int_A f d\mu$ for all $A \in \mathcal{A}$.

Now we suppose that μ is σ-finite. There exist $F_1 \subseteq F_2 \subseteq \cdots$ such that $\bigcup_{i=1}^{\infty} F_i = X$ and $\mu(F_i) < \infty$. Let μ_i be the restriction of μ to F_i. Define ν_i by the restriction of ν to F_i. If $\mu(A) = 0$, then $\mu(A \cap F_i) = 0$, hence $\nu(A \cap F_i) = 0$, and thus $\nu_i(A) = 0$. Therefore $\nu_i \ll \mu_i$. If f_i is the function that we found above, we must have that $f_i = f_j$ on F_i if $i \leq j$. Define f by $f(x) = f_i(x)$ if $x \in F_i$. Then for each $A \in \mathcal{A}$

$$\nu(A \cap F_i) = \nu_i(A) = \int_A f_i d\mu_i = \int_{A \cap F_i} f d\mu.$$

Letting $i \to \infty$ shows that f is the desired function. □

Chapter 6

Topology

Here we introduce the notion of a topological space along with many examples. This section is based off of [AF08], a much more detailed discussion can be found there with lots of further examples and concepts.

Definition 61. Let X be a non-empty set and let $\mathcal{T}$ be a collection of subsets of X such that

(a) $X \in \mathcal{T}$ and $\emptyset \in \mathcal{T}$.

(b) If $\{E_\alpha\}_{\alpha \in I}$ is a arbitrary collection of sets in $\mathcal{T}$, then $\bigcup_{\alpha \in I} E_i \in \mathcal{T}$.

(c) If $E_1, \ldots, E_n$ are each in $\mathcal{T}$, then $\bigcap_{i=1}^{n} E_i \in \mathcal{T}$.

Definition 62. The elements of $\mathcal{T}$ are called the *open sets*.

If a space is a metric space, then it is automatically a metric space by letting $\mathcal{T}$ be the collection of open sets. Now we give many examples of topologies.

Example. (a) The simplest topology on a space X is the topology $\mathcal{T} = \{\emptyset, X\}$. This is called the *trivial topology*. In a sense that we will soon define, it is the coarsest topology we can put on a space. It does not give much information about the space.

(b) In the opposite direction, for a set X, we can take $\mathcal{T}$ to be the collection of every subset of X. We call this the *discrete topology*. This is a very, in a sense that we will soon define, fine topology. It carries every detail of the space. Often times neither the trivial topology nor the discrete topology is very useful, since the discrete topology does not carry enough information about the space, while the discrete topology carries too much space. Something in the middle can be more useful, but the trivial and discrete topologies are good basic examples to exhibit some topological phenomena.

(c) The *Sierpinski topology* on a set $X = \{a, b\}$ is the set $\mathcal{T} = \{\emptyset, X, \{b\}\}$.

(d) Let $X = \mathbb{R}$, then letting $\mathcal{T}$ be the collection of open sets is a topology. In this case an open set will be a union of sets of the form (a, b) with $a, b \in \mathbb{R}$. So $(1, 2) \cup (\pi, 5)$ is an open set. As is $\bigcup_{n=1}^{\infty} (n, n+1)$. This is called the *Euclidean* topology on $\mathbb{R}$.

(e) Similarly, we can take $X = \mathbb{R}^d$ and let $\mathcal{T}$ be the open sets in X. This time they will be unions of open balls:

$$B(x, r) = \left\{ y \in \mathbb{R}^d \mid d(x, y) < r \right\}.$$

This is also called the *Euclidean topology on* $\mathbb{R}^d$.

(f) We can put topologies on finite sets as well. Let $X = \{1, 2, 3, 4\}$ and let

$$\mathcal{T} = \{\emptyset, \{2\}, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}, \{1, 2, 3, 4\}\}.$$

The reader can check that $\mathcal{T}$ satisfies the axioms of a topology.

(g) Let X be an infinite set, we can take $X = \mathbb{R}$ for example. Then let

$$\mathcal{T} = \{A \subseteq X \mid A = \emptyset \text{ or } X \backslash A \text{ is finite}\}.$$

This is called the *cofinite topology*.

(h) This example comes from algebraic geometry. Take $X = \mathbb{R}^2$, and consider polynomials with coefficients in $\mathbb{R}$ over two variables, we call this set $\mathbb{R}[X, Y]$ where X and Y are the two variables. This set contains elements like $X^2 + Y^3, X^7, 3, 4 + XY^2 + Y^6X$, and so on. Then for any subset $S \subset \mathbb{R}[X, Y]$, let

$$V(S) = \left\{ (x, y) \in \mathbb{R}^2 \mid f(x, y) = 0 \text{ for all } f \in S \right\} \subseteq \mathbb{R}^2.$$

Then let $\mathcal{T}$ be sets of the form $R^2 \backslash V(S)$ for any subset $S \subset \mathbb{R}[X, Y]$. This is called the *Zariski topology* and is crucial in algebraic geometry.

(i) The above example can be extended for those who are familiar with a little ring theory. Let R be any commutative ring, meaning multiplication is commutative, i.e., $ab = ba$ for any $a, b \in R$. Then let $\operatorname{Spec} R$ be the set of all prime ideals in R, and take an ideal $I \triangleleft R$, and consider the set

$$V(I) = \{\mathfrak{p} \in \operatorname{Spec} R \mid I \subseteq \mathfrak{p}\}$$

the set of all prime ideals containing the ideal I. Then if we let $\mathcal{T}$ contain sets of the form $\operatorname{Spec} A \backslash V(I)$ for some ideal $I \triangleleft A$, this forms a topology also called the *Zariski topology*. These two topologies might look very different, but over a single variable polynomial ring over an algebraically closed field they are actually identical due to the weak nullstellensatz. This second version of the Zariski topology is just a generalized version of the first that is useful for scheme theory. This example is meant to show that topologies can show up everywhere.

This list should show that there is not a unique topology on a space. Every space has at least two, the discrete topology and the trivial topology. On $\mathbb{R}$ we demonstrated four different examples, the discrete, the trivial, the Euclidean and the cofinite. There are many others that can be put on $\mathbb{R}$ as well.

Definition 63. Let X be a set and let $\mathcal{T}_1, \mathcal{T}_2$ be two topologies on X. If $\mathcal{T}_1 \subset \mathcal{T}_2$, then $\mathcal{T}_2$ is said to be *finer* than $\mathcal{T}_1$, and $\mathcal{T}_1$ is said to be coarser than $\mathcal{T}_2$. Furthermore, if $\mathcal{T}_2$ is finer than $\mathcal{T}_1$ but not equal to $\mathcal{T}_1$, then $\mathcal{T}_2$ is said to be *strictly finer* than $\mathcal{T}_1$. *Strictly coarser* is defined analogously.

Example. For any set X, the trivial topology is coarser than the discrete topology and the discrete topology is finer than the trivial topology. For any nonempty set ,the discrete topology is strictly finer than the trivial topology and the trivial topology is strictly coarser than the discrete topology.

We quickly define the notion of compactness which will be quite useful later on.

Definition 64. An *open cover* of a set E in a metric space X is a collection $\{G_\alpha\}$ of open subsets of X such that $E \subset \bigcup_\alpha G_\alpha$.

Definition 65. A subset K of a metric space X is said to be *compact* if every open cover of K contains a *finite* subcover. This means that if $\{G_\alpha\}$ is an open cover of K, then there are finitely many indices $\alpha_1, \dots, \alpha_n$ such that

$$K \subset G_{\alpha_1} \cup \cdots \cup G_{\alpha_n}.$$

Theorem 66 ([Rud76] Theorem 2.36)**.** *If $\{K_\alpha\}$ is a collection of compact subsets of a metric space X such that the intersection of every finite subcollection of $\{K_\alpha\}$ is nonempty, then $\bigcap_\alpha K_\alpha$ is nonempty.*

Proof. Fix K_1 of $\{K_\alpha\}$, and let $G_\alpha = K_\alpha^c$. Assume that no point of K_1 belongs to every K_α. Then the G_α form an open cover of K_1, and since K_1 is compact, there are finitely many $G_{\alpha_1}, \dots, G_{\alpha_n}$ such that $K_1 \subset G_{\alpha_1} \cup \cdots G_{\alpha_n}$. But this means that

$$K_1 \cap K_{\alpha_1} \cap \cdots \cap K_{\alpha_n} = \emptyset,$$

a contradiction. □

Chapter 7

Vector Spaces

We give a brief introduction to vector spaces, mainly with just basic definitions, based off of [Lan87]

Definition 67. A field K is a set with two binary operations $+$ and $\cdot$, along with two elements 0, the additive identity, and 1, the multiplicative identity, meaning $0 + x = x + 0 = x$ and $1 \cdot x = x \cdot 1 = x$ satisfying

(a) If $x, y \in K$, then $x + y \in K$ and $x \cdot y \in K$.

(b) If $x \in K$, then there is an element $-x \in K$ such that $x + (-x) = 0$, we call $-x$ the *additive identity* of x. If $x \neq 0$, then there is an element x^{-1} such that $x \cdot x^{-1} = 1$, we call x^{-1}, the *multiplicative identity* of x.

(c) The elements 0 and 1 are in K.

Example. Many of the objects we are most familiar with in math are fields.

(a) The real numbers, $\mathbb{R}$ are a field.

(b) The complex numbers $\mathbb{C}$ are a field.

(c) The rational number $\mathbb{Q}$ are a field.

(d) The integers, $\mathbb{Z}$, are not a field. This is because not every element has a multiplicative inverse, for example there is no integer x such that $x \cdot 2 = 1$.

(e) There are other types of fields, for example, the set

$$\mathbb{Q}(\sqrt{2}) = \left\{ a + b\sqrt{2} \mid a, b \in \mathbb{Q} \right\}$$

is a field.

(f) The set

$$\mathbb{R}(x) = \left\{ \frac{f(x)}{g(x)} \mid f(x), g(x) \text{are polynomials with coefficients in } \mathbb{R} \text{ and } g(x) \neq 0 \right\}$$

is a field as well.

Definition 68. A vector space V over a field K is a set of objects which satisfies the following properties:

(a) Given elements $u, v, w \in V$, we have
$$(u+v)+w = u+(v+w).$$

(b) There is an element of V, denoted 0, such that
$$0+u = u+0 = u$$
for all $u \in V$.

(c) Given an element u of V, there exists an element $-u$ in V such that
$$u+(-u) = 0.$$

(d) For all elements u, v of V, we have
$$u+v = v+u.$$

(e) If $c \in K$, then $c(u+v) = cu+cv$.

(f) If $a, b \in K$, then $(a+b)v = av+bv$.

(g) If $a, b \in K$, then $(ab)v = a(bv)$.

(h) For all elements u of V, we have $1 \cdot u = u$.

Example. We are also familiar with many examples of vector spaces.

(a) $\mathbb{R}$ is a vector space over $\mathbb{R}$.

(b) $\mathbb{R}^d$ is a vector space over $\mathbb{R}$ for any $d > 0$.

(c) Similarly, $\mathbb{C}^d$ is a vector space over $\mathbb{C}$.

(d) The field we looked at above, $\mathbb{Q}(\sqrt{2})$ is a vector space over $\mathbb{Q}$.

(e) Another example is the set of continuous functions $f\colon [a,b] \to \mathbb{R}$, denoted $C([a,b],\mathbb{R})$. This is a vector space over $\mathbb{R}$, as we can add two continuous functions and get a continuous function, and multiplying a continuous function by an element in $\mathbb{R}$ is still a continuous function. Also, the constant 0 function is a continuous function as is the function that is constant 1.

Chapter 8

Metric Spaces

More information on this section can be found in [Rud76]

Definition 69. We say that (X, d), where X is a set and d is a function $d\colon X \times X \to \mathbb{R}$, is a *metric space*, if for any two points $p, q \in X$, d satisfies the following:

(a) $d(p, q) > 0$ if $p \neq q$,and $d(p, p) = 0$.

(b) (symmetry) $d(p, q) = d(q, p)$.

(c) (triangle inequality) $d(p, q) \leq d(p, r) + d(r, q)$ for any $r \in X$.

Any function with these properties is called a *metric*.

Remark 70. The third property, the triangle inequality, says that it should be shorter to travel in a straight line between two points than to first go to another, intermediate point, and then to the destination. In figure 8.1, the distance from p to q is shorter than the distance from p to r plus the distance from r to q.

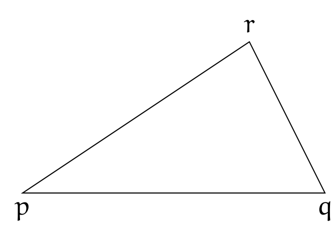


Figure 8.1: Illustration of the triangle inequality

Example.

(a) Let $X = \mathbb{R}$. Then the absolute value function is a metric on $\mathbb{R}$, and thus $(\mathbb{R}, |\cdot|)$ is a metric space. Indeed, $|x - y| = 0$ if and only if $x = 0$, otherwise it is greater than 0. Similarly, $|x - y| = |y - x|$, and finally, the absolute value satisfies the triangle inequality: $|x - y| \leq |x - z| + |z - y|$.

(b) Let $X = \mathbb{R}^2$ and let $d(\vec{x}, \vec{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$ where $\vec{x} = (x_1, x_2)$ and $\vec{y} = (y_1, y_2)$. This satisfies the properties of a metric, and so $(\mathbb{R}^2, d)$ is a metric space. This is called the *Euclidean metric*.

(c) Let $X = \mathbb{R}^2$ and let $d(\vec{x}, \vec{y}) = |x_1 - y_1| + |x_2 - y_2|$. This is a different metric, but we will soon see that these two metrics on $\mathbb{R}^2$ are equivalent. This metric is called the *taxicab metric*.

(d) Let $X = \mathbb{R}^n$, and let $p \in \mathbb{R}$ and $p \geq 1$. Then

$$d_p(\vec{x}, \vec{y}) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

is a metric on $\mathbb{R}^n$, setting $n = 1$ and $p = 1$ we get the absolute value on $\mathbb{R}$, and letting $n = 2$ and $p = 1$ gives us the taxi cab metric, and setting $n = 2$ and $p = 2$ gives the Euclidean metric. These are called the d^p metrics on $\mathbb{R}^n$.

(e) Let $X = \mathbb{R}^n$ and let $d \colon X \times X \to \mathbb{R}$ be defined as

$$d(\vec{x}, \vec{y}) = \sup \{ |x_i - y_i| \mid i = 1, 2 \ldots, n \} .$$

This is another metric on $\mathbb{R}^n$, called the d_∞ metric.

(f) Let $X = C([0,1], \mathbb{R})$ be the set of continuous functions from $[0,1]$ to $\mathbb{R}$. Then

$$d(f, g) = \sup \{ d(f(x), g(x)) \mid x \in [0,1] \}$$

is a metric on X, often called the *supremum metric.*

(g) Another example on $\mathbb{R}$ is to take

$$d(x, y) = |\arctan(x) - \arctan(y)| .$$

Definition 71. Let (X, d) be a metric space.

(a) The *open ball centered at* x *of radius* r is $\mathbb{B}(x, r) = \{ y \in X \mid d(x, y) < r \}$.

(b) The *closed ball centered at* x *of radius* r is $\overline{\mathbb{B}}(x, r) = \{ y \in X \mid d(x, y) \leq r \}$.

(c) The *sphere centered at* x *of radius* r is $S(x, r) = \{ y \in X \mid d(x, y) = r \}$.

Example. Consider the metric space $(\mathbb{R}, |\cdot|)$, the real line with the usual absolute value. Then $\mathbb{B}(0,1), \overline{\mathbb{B}}(0,1)$, and $S(0,1)$ are depicted in figure 8.2. The open ball is $\mathbb{B}(0,1) = \{ x \in \mathbb{R} \mid |x| < 1 \}$, the closed ball is $\overline{\mathbb{B}}(0,1) = \{ x \in \mathbb{R} \mid |x| \leq 1 \}$, and the sphere is $S(0,1) = \{ x \in \mathbb{R} \mid |x| = 1 \} = \{ -1, 1 \}$.

Example. Figure 8.3 shows $\mathbb{B}((0,0), 1)$ in $\mathbb{R}^2$ under the ℓ^1, ℓ^2, and ℓ^∞ metrics.

Definition 72. Let (X, d) be a metric space and let $Y \subset X$. We say that

(a) A point $y \in Y$ is an *interior point* of Y if $\mathbb{B}(y, r) \subset Y$ for some $r > 0$. We call the set of interior points the *interior* of Y, and denote it by $\mathrm{int}(M)$.

(b) a point $y \in Y$ is an *exterior point* if $\mathbb{B}(y, r) \cap Y = \emptyset$ for some $r > 0$. We call the set of exterior points the *exterior* of Y, and denote it by $\mathrm{ext}(M)$.

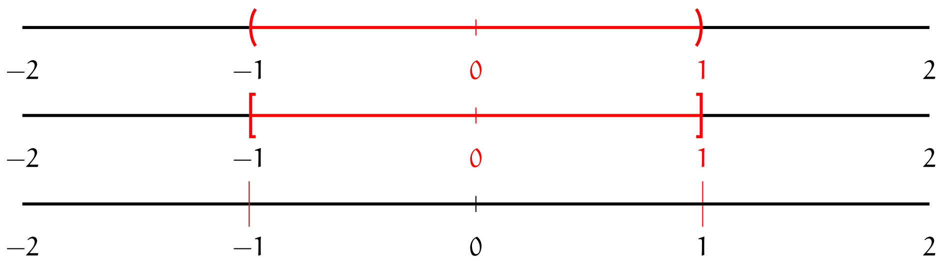


Figure 8.2: $\mathbb{B}(0, 1)$, $\overline{\mathbb{B}}(0, 1)$, and $S(0, 1)$.

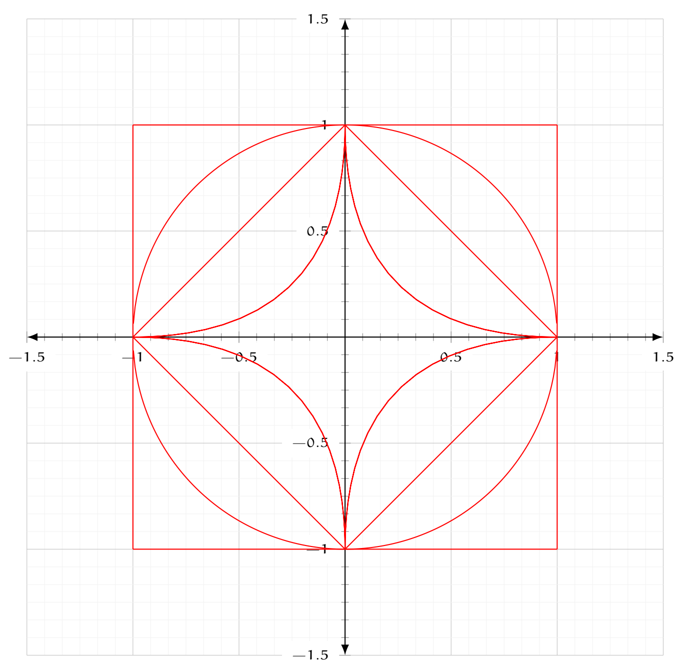


Figure 8.3: $\mathbb{B}((0, 0), 1)$ in $\mathbb{R}^2$ under the $d_{1/2}$, d_1, d_2, and d_∞ metrics.

(c) a point $y \in Y$ is a *boundary point* of Y if for every $r > 0$, $\mathbb{B}(y,r) \cap M \neq \emptyset$ and $\mathbb{B}(y,r) \cap M^c \neq \emptyset$. We call the set of boundary points the *boundary* of Y, and denote it by ∂Y.

Remark 73. For every $Y \subset X$, we can partition X into three pieces:

$$X = \operatorname{int}(Y) \cup \operatorname{ext}(Y) \cup \partial(Y).$$

Example.

(a) $\operatorname{int}(\mathbb{B}(y,r)) = \{x \in X \mid d(y,x) < r\} = \operatorname{int}(\overline{\mathbb{B}}(y,r))$.

(b) $\operatorname{ext}(\mathbb{B}(y,r)) = \{x \in X \mid d(y,x) \leq r\} = \operatorname{ext}(\overline{\mathbb{B}}(y,r))$.

(c) $\partial(\mathbb{B}(y,r)) = \{x \in X \mid d(y,x) = r\} = \partial(\overline{\mathbb{B}}(y,r))$.

Definition 74. For $Y \subset X$, its *closure* is $\overline{Y} = \operatorname{int}(Y) \cup \partial(Y)$.

Example. The closure of $\mathbb{B}(y,r)$ is $\overline{\mathbb{B}}(y,r)$, that is,

$$\overline{\mathbb{B}(y,r)} = \overline{\mathbb{B}}(y,r).$$

Proposition 75. *Let (X,d) be a metric space and let $Y \subset X$. Then a point $x \in X$ belongs to $\overline{Y}$ if and only if for every $r > 0$, $\mathbb{B}(x,r) \cap Y \neq \emptyset$.*

The operations int, ext, and ∂ are not independent of each other.

Proposition 76. *Let (X,d) be a metric space and let $Y \subset X$, then*

(a) $\operatorname{ext}(Y) = \operatorname{int}(Y^c)$.

(b) $\partial(Y) = \partial(Y^c)$.

(c) $(\int(Y))^c = \overline{Y^c}$.

Definition 77. Let (X,d) be a metric space, and let $Y \subset X$. Then

(a) Y is *open* if $Y \cap \partial Y = \emptyset$.

(b) Y is *closed* if $\partial Y \subset Y$.

Example.

(a) $\mathbb{B}(y,r)$ is open.

(b) $\overline{\mathbb{B}}(y,r)$ is closed.

Remark 78.

(a) A subset Y of a metric space (X,d) is open if and only if $\operatorname{int}(Y) = Y$, and it is closed if and only if $\overline{Y} = Y$.

(b) The topology of a metric space, (X,d) is the set of open subsets of X.

(c) Y is closed if and only if Y^c is open and Y is open if and only if Y^c is closed.

Proposition 79. *Let* (X, d) *be a metric space.*

(a) The union of an arbitrary family of open subsets of X *is open.*

(b) The intersection of an arbitrary finite family of open subsets of X *is open.*

(c) The union of an arbitrary finite family of closed subsets of X *is closed.*

(d) The intersection of an arbitrary family of closed subsets of X *is closed.*

Proof.

(a) Let us consider an arbitrary family of open sets $\{U_\lambda \mid \lambda \in \Lambda\}$. We want to show that $\bigcup_{\lambda\in\Lambda} U_\lambda$ is open. Let x be a point of this union. That is, $z \in U_{\lambda_0}$ for some $\lambda_0 \in \Lambda$. As U_{λ_0} is open, $x \in \text{int}(U_{\lambda_0})$, meaning that there exists $r > 0$ such that $\mathbb{B}(x, r) \subset U_{\lambda_0}$. But then
$$\mathbb{B}(x, r) \subset \bigcup_{\lambda\in\Lambda} U_{\lambda_0}.$$
Showing that x is an interior point of the union.

(b) Let us consider an arbitrary finite family of open sets $\{U_n \mid n \in \{1, \ldots, N\}\}$. We want to show that $\bigcap_{n=1}^{N} U_n$ is open. That is, $x \in U_n$ for every $n \in \{1, \ldots, N\}$. Then as U_n is open, there exists a radius $r_n > 0$ such that $\mathbb{B}(x, r_n) \subset U_n$. Let us define $r = \min\{r_n \mid n \in \{1, \ldots, N\}\}$. This agrees with one of the r_n's so it is positive. Further, for every $n \in \{1, \ldots, N\}$, we have that
$$\mathbb{B}(x, r) \subset \mathbb{B}(x, r_n) \subset U_n.$$
That is $\mathbb{B}(x, r) \subset \bigcap_{n=1}^{N} U_n$, so we are done.

(c) Let us consider an arbitrary finite family of closed set $\{V_n \mid n \in \{1, \ldots, N\}\}$. We want to show that $\bigcup_{n=1}^{n} V_n$ is closed. Equivalently,
$$\left(\bigcup_{n=1}^{N} V_n\right)^c$$
is open. The above is just $\bigcap_{n=1}^{N} V_n^c$, where each V_n^c is open. So we are done by part (2).

(d) Let us consider an arbitrary family of closed sets $\{V_\lambda \mid \lambda \in \Lambda\}$. We want to show that $\bigcap_{\lambda\in\Lambda} V_\lambda$ is closed. Equivalently, $(\bigcap_{\lambda\in\Lambda} U_\lambda)^c$ is open. This is $\bigcup_{\lambda\in\Lambda} V_\lambda^c$ where each V_λ^c is open. Thus we are done by part (1).

□

Definition 80. A topological space X is *Hausdorff* if for every pair of distinct points $x, y \in X$, there exist disjoint open sets U and V such that $x \in U$ and $y \in V$.

Example.

(a) Take the topological space $\mathbb{R}$ with the Euclidean topology. Recall that this topology arises from the usual absolute value on $\mathbb{R}$. Let $x, y \in \mathbb{R}$, $x \neq y$ and let $r = |x - y|$. Then the open sets $\mathbb{B}(x, r/2)$ and $\mathbb{B}(y, r/2)$ are disjoint open sets containing x and y. This can be seen in figure 8.4.

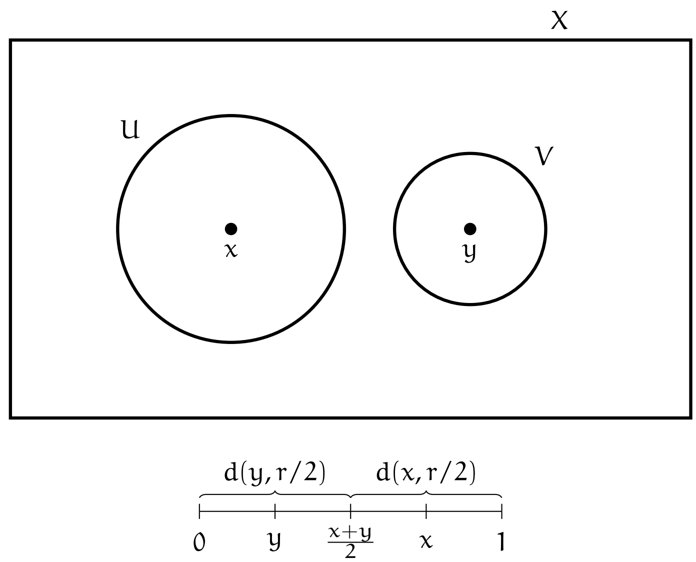


Figure 8.4: $\mathbb{R}$ with the Euclidean topology is Hausdorff.

(b) More generally, for any metric space (X, d), let $x, y \in X$ and let $r = d(x, y)$, then $\mathbb{B}(x, r/2)$ and $\mathbb{B}(y, r/2)$ are two disjoint open sets containing x and y, and thus X is Hausdorff.

(c) An example of a non-Hausdorff space is $\mathbb{R}$ under the cofinite topology. Let $x, y \in \mathbb{R}$. Any open set U containing x, has a finite complement, and any open set V containing y has a finite complement. Thus since both U and V are infinite sets with finite complements, they must intersect at some point, so this space cannot be Hausdorff.

(d) Other examples of non-Hausdorff spaces are $\mathbb{C}^n$ under the Zariski topology, or Spec A under the Zariski topology for any commutative ring A.

Chapter 9

Normed Spaces

The references used for this section are [Con90] and [Kre78].

Definition 81. If X is a vector space over a field K, a *seminorm* is a function $p\colon X \to [0,\infty)$ having the properties:

(a) $p(x+y) \leq p(x) + p(y)$ for all $x, y \in X$.

(b) $p(\alpha x) = |\alpha|\, p(x)$ for all $\alpha \in K$ and $x \in X$.

It follows that $p(0) = p(0\cdot 0) = 0 \cdot p(0) = 0$. A *norm* is a seminorm p such that $x = 0$ if $p(x) = 0$. Usually we will denote a norm by $\|\cdot\|$.

Definition 82. A *normed space* is a pair $(X, \|\cdot\|)$ where X is a vector space and $\|\cdot\|$ is a norm.

Definition 83. If p and q are positive real numbers such that

$$\frac{1}{p} + \frac{1}{q} = 1,$$

then we call p and q a pair of *Hölder conjugate exponents*.

Example. Letting $p = q = 2$, we get a pair of conjugate exponents. We take $p = 1, q = \infty$ to be a pair of conjugate exponents.

First we prove two famous inequalities that will help us establish some examples of normed spaces.

Definition 84. A real function f defined on an interval (a, b) where $-\infty \leq a < b \leq \infty$, is called *convex* if the inequality

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$$

holds whenever $a < x, y < b$ and $0 \leq \lambda \leq 1$. On the other hand, a function is *concave* if the reverse inequality holds:

$$f((1-\lambda)x + \lambda y) \geq (1-\lambda)f(x) + \lambda f(y)$$

Remark 85. A function is convex if $x < t < y$, the point $(t, f(t))$ should lie below the line connecting $(x, f(x))$ and $(y, f(y))$. This can be seen in figure 9.1. On the other hand the function $\sqrt{x}$ is not convex on $[0,1]$. Indeed, let $x = \frac{1}{2}, y = \frac{3}{4}$ and $\lambda = \frac{3}{4}$, then

$$\sqrt{\frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\frac{3}{4}} = 0.75 \geq \frac{3}{4}\sqrt{\frac{1}{2}} + \frac{1}{4}\sqrt{\frac{3}{4}} \approx 0.655.$$

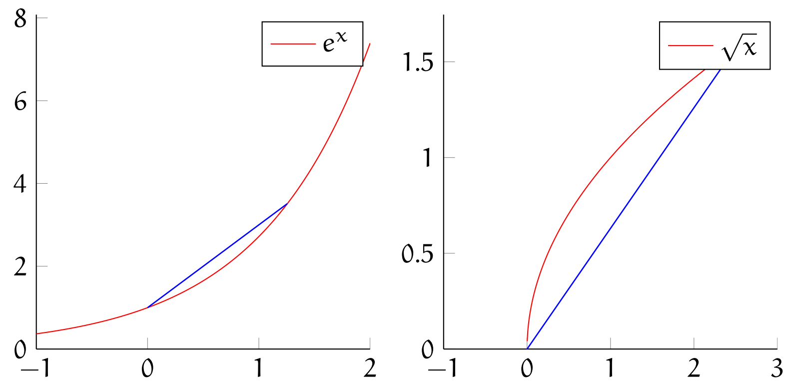


Figure 9.1: The function e^x is convex while $\sqrt{x}$ is not

We now turn to some of the most important spaces in functional analysis.

Definition 86. Define $\mathcal{L}^p(X) = \left\{f\colon X \to \mathbb{R} \mid \int_X |f(x)|^p \, d\mu < \infty\right\}$ where μ is the Lebesgue measure.

Theorem 87 (Hölder's Inequality, [Rud87] Theorem 3.5)**.** *Let* p *and* q *be conjugate exponents,* $1 < p < \infty$. *Let* $f, g \in \mathcal{L}^p(X)$. *Then*

$$\int_X f(x)g(x)\,d\mu \leq \left\{\int_X f(x)^p \, d\mu\right\}^{1/p} \left\{\int_X g(x)^q \, d\mu\right\}^{1/q}. \tag{9.1}$$

Proof. Let A and B be the two factors on the right of (9.1). If $A = 0$, then $f = 0$ hence $fg = 0$, so (9.1) holds. If $A > 0$ and $B = \infty$, the statement is again trivial. So we only need to consider the case $0 < A < \infty$, $0 < B < \infty$. Put

$$F = \frac{f}{A}, \ G = \frac{g}{B}.$$

This gives

$$\int_X F^p \, d\mu = \int_X G^q \, d\mu = 1. \tag{9.2}$$

Indeed, we have

$$\int_X \left(\frac{f(x)}{A}\right)^p d\mu = \frac{f(x)^p}{\int_X f(x)^p \, d\mu} = \frac{1}{\int_X f(x)\,d\mu} \int_X f(x)\,d\mu = 1,$$

since the integral $\int_X f(x)\,d\mu$ is just a constant we can move it outside the integral. If $c \in [a, b]$ such that $0 < F(c) < \infty$ and $0 < G(c) < \infty$, there are real numbers s and t such that $F(c) = e^{s/p}$, $G(c) = e^{t/q}$. The existence of s and t come from the fact that the function $\left(e^{1/p}\right)^x$ and $\left(e^{1/q}\right)^x$ are surjective on $(0, \infty)$. Since $\frac{1}{p} + \frac{1}{q} = 1$, and e^x is convex, we have

$$e^{s/p + t/q} \leq \frac{e^s}{p} + \frac{e^t}{q}.$$

This applies the convex inequality

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$$

whenever $-\infty < x, y < \infty$ and $0 \le \lambda \le 1$. Plugging in $x = s, y = t, \phi(x) = e^x$, and $\lambda = \frac{1}{p}$ we get the above inequality. It follows that

$$F(c)G(c) \le \frac{F(c)^p}{p^{-1}} + \frac{G(c)^p}{q} \tag{9.3}$$

for every $c \in [a, b]$. Integrating 9.3 yields

$$\int_a^b F(x)G(x)\,dx \le \frac{1}{p} + \frac{1}{q} = 1$$

by 9.2. Then inserting $F = \frac{f}{A}$, $G = \frac{g}{B}$, into 9.3, we get

$$\frac{\int_a^b f(x)g(x)\,dx}{\left(\int_a^b f(x)^p\,dx\right)^{1/p}\left(\int_a^b g(x)^q\right)^{1/q}} \le 1,$$

giving us the inequality. □

Remark 88. In $\mathbb{R}^n$, Hölder's inequality looks like

$$\sum_{i=1}^n |x_i||y_i| \le \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}\left(\sum_{i=1}^n |y_i|^q\right)^{1/q}$$

where $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$.

Theorem 89 (Minkowski's Inequality, [Rud87] Theorem 3.5)**.** *Let* $1 < p < \infty$. *Let* $C[a, b]$ *be the space of continuous functions* $f\colon [a, b] \to \mathbb{R}$. *Let* $f, g \in C[a, b]$. *Then*

$$\left(\int_a^b (f(x) + g(x))^p\,dx\right)^{1/p} \le \left(\int_a^b f(x)^p\,dx\right)^{1/p} + \left(\int_a^b g(x)^p\,dx\right)^{1/p}. \tag{9.4}$$

Proof. First we let q be the conjugate of p. Then we write

$$(f + g)^p = f \cdot (f + g)^{p-1} + g \cdot (f + g)^{p-1}. \tag{9.5}$$

We want to apply Hölder's inequality, but to do this we need to make sure that $(f + g)^{p-1} \in \mathcal{L}^q$, i.e. we need to check that

$$\int_X ((f + g)^{p-1})^q\,d\mu < \infty.$$

Indeed,

$$\int_X ((f + q)^{p-1})^q\,d\mu = \int_X (f + g)^{q(p-1)}\,d\mu = \int_X (f + q)^p\,d\mu < \infty.$$

Hölder's inequality gives

$$\int_X f(x) \cdot (f + g)^{p-1}\,d\mu \le \left(\int_X f(x)^p\,d\mu\right)^{1/p}\left(\int_X (f(x) + g(x))^{(p-1)q}\,d\mu\right)^{1/q}. \tag{9.6}$$

We have an equivalent inequality with f switched with g. Adding those two together and noting that $(p-1)q = p$ since p and q are conjugates, so $p+q = pq$, we get

$$\int_X (f(x)+g(x))^p \, d\mu \le \left(\int_X (f(x)+g(x))^p \, d\mu\right)^{1/q} \left[\left(\int_X f(x)^p \, d\mu\right)^{1/p} + \left(\int_X g(x)^p \, d\mu\right)^{1/p}\right]. \tag{9.7}$$

It is enough to prove the inequality in the case that the left hand side is greater than 0 and the right hand side is less than ∞, since otherwise it is trivial. The convexity of the function t^p for $0 < t < \infty$ shows that

$$\left(\frac{f+g}{2}\right)^p \le \frac{1}{2}\left(f^p + g^p\right).$$

Hence the left hand side of (9.7) is less than ∞ and the inequality follows from (9.7) if we divide by the first factor on the right hand side of (9.7), since $1 - \frac{1}{q} = \frac{1}{p}$ since we have

$$\frac{\int_X (f(x)+g(x))^p \, d\mu}{\left(\int_X (f(x)+g(x))^p \, d\mu\right)^{1/q}} = \left(\int_X (f(x)+g(x))^p \, d\mu\right)^{1/p} \le \left(\int_X f(x)^p \, d\mu\right)^{1/p} + \left(\int_X g(x)^p \, d\mu\right)^{1/p}.$$

□

These two theorems are so important that we provide an alternate, more geometric proof. We first need a lemma.

Lemma 90 (Young's Inequality, [Roy88] page 140)**.** *Assume $1 < p, q < \infty$ are such that $\frac{1}{p} + \frac{1}{q} = 1$. Then for any $x, y \ge 0$. we have*

$$xy \le \frac{x^p}{p} + \frac{y^q}{q}.$$

Proof. We first claim that if $0 \le s \le x$ and $0 \le t \le y$, then $t \ge s^{p-1}$ or $s \ge t^{q-1}$. Indeed, if $t < s^{p-1}$, then raising both sides to the power $q-1$ we get $t^{q-1} < s$, since $(p-1)(q-1) = pq - p - q + 1 = 1$ which comes from the fact that $pq = p + q$ since p and q are Hölder conjugates. Then the rectangle $[0, x] \times [0, y]$ are covered by the sets

$$\left\{(s,t) \mid 0 \le s \le x, s^{p-1} \le t \le 1\right\}, \quad \left\{(s,t) \mid 0 \le t \le y, t^{q-1} \le s \le 1\right\}.$$

The area of the rectangle is xy, while those of the covering sets are

$$\int_0^x s^{p-1} dt = \frac{x^p}{q}, \quad \int_0^y t^{q-1} dt = \frac{y^q}{q}.$$

Thus we get the desired result. □

Now we can prove Hölder's inequality.

Alternate Proof of Hölder's Inequality. If $\left(\int_X |f(x)|^p \, d\mu\right)^{1/p} = 0$ or $\left(\int_X |g(x)|^q \, d\mu\right)^{1/q} = 0$, then $f = 0$ almost everywhere, or $g = 0$ almost everywhere, so $fg = 0$ almost everywhere, therefore we may assume, after a suitable normalization, that $\left(\int_X |f(x)|^p \, d\mu\right)^{1/p} = \left(\int_X |g(x)|^q \, d\mu\right)^{1/q} = 1$.

For any $x \in X$, we apply Young's inequality to see that

$$|f(x)g(x)| \le \frac{|f(x)|^p}{p} + \frac{|g(x)^q|}{q}.$$

Integrating both sides we get

$$\int_X |fg|\, d\mu \le \frac{\left(\int_X |f(x)|^p\, d\mu\right)^{1/p}}{p} + \frac{\left(\int_X |g(x)|^q\, d\mu\right)^{1/q}}{q} = \frac{1}{p} + \frac{1}{q} = 1,$$

and we get the desired result. □

Example.

(a) Take $X = \mathbb{R}^n$ and let $p \ge 1$ with $p \in \mathbb{R}$

$$\|\vec{x}\| = \left(\sum_{i=1}^{n} |\vec{x}|^p\right)^{1/p}.$$

Notice that this is very similar to the metrics we defined earlier, this is because the metric actually arises from the norm, if we set

$$d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|$$

we will get the definition of the ℓ^p metric we had earlier. Not all metrics arise from norms however.

(b) An example of a metric space that does not arise from a norm is the sequence space s. The space consists of the set of all sequences of complex numbers, and the metric is defined as

$$d(x, y) = \sum_{j=1}^{\infty} \frac{1}{2^j} \frac{|\xi_j - \eta_j|}{1 + |\xi_j - \eta_j|}$$

where $x = (\xi_j)$ and $y = (\eta_j)$. It can be seen from lemma 92 that this metric cannot arise from a norm.

(c) Consider $X = C([a, b], \mathbb{R})$ the space of continuous functions from $[a, b]$ to $\mathbb{R}$. This is a normed space with norm

$$\|f\|_\infty = \max\{|f(x)| \mid x \in [a, b]\}.$$

Now we show that the norm $\|\cdot\|_\infty$ is indeed a norm. We establish the three conditions:

(a) Let $f, g \in C([a, b], \mathbb{R})$, and consider

$$\|f + g\|_\infty = \max\{|f(x) + g(x)| \mid x \in [a, b]\} \le \max\{|f(x)| \mid x \in [a, b]\} + \max\{|g(x)| \mid x \in [a, b]\} = \|f\|_\infty + \|g\|_\infty,$$

where the middle inequality comes from the fact that f and g might take their maximum values at different x values.

(b) Let $\alpha \in \mathbb{R}$ and consider

$$\|\alpha f\|_\infty = \max\{|\alpha f(x)| \mid x \in [a, b]\} = \max\{|\alpha|\,|f(x)| \mid x \in [a, b]\} = |\alpha| \max\{|f(x)| \mid x \in [a, b]\}.$$

(c) Suppose $\|f\|_\infty = 0$, this means that $\max\{|f(x)| \mid x \in [a, b]\} = 0$, and thus $f = 0$.

(d) Let $X = C([a, b], \mathbb{R})$, another norm we can put on this space using the Riemann integral, we could also use the Lebesgue integral, but it is unnecessary since we are only integrating continuous functions. Let

$$\|x\|_1 = \int_0^1 |x(t)|\, dt.$$

(e) Another norm on continuous functions $C([a,b],\mathbb{R})$ is defined as

$$\|f\|_2 = \left(\int_a^b |f(x)|^2\, dx\right)^{1/2}.$$

More generally, for any $p \geq 1, p \in \mathbb{R}$, there is a norm defined as

$$\|f\|_p = \left(\int_a^b |f(x)|^p\, dx\right)^{1/p}.$$

We quickly prove that $\|\vec{x}\|_p$ is indeed a norm. We verify the three conditions:

(a) Let $f, g \in C([a,b],\mathbb{R})$, then

$$\|f+g\|_p = \left(\int_a^b |f(x)+g(x)|^p\, dx\right)^{1/p} \leq \left\{\int_a^b |f(x)|^p\, dx\right\}^{1/p} + \left\{\int_a^b |g(x)|^p\, dx\right\}^{1/p}$$

by Minkowski's inequality.

(b) Let $\alpha \in \mathbb{R}$ and consider

$$\|\alpha f\|_p = \left(\int_a^b |\alpha f(x)|^p\, dx\right)^{1/p} = \left(|\alpha|^p \int_a^b |f(x)|^p\, dx\right)^{1/p} = |\alpha|\, \|f\|_p\,.$$

(c) Suppose $\|f\|_p = 0$. Then

$$\begin{aligned} \|f\|_p = \left(\int_a^b |f(x)|^p\, dx\right)^{1/p} &= 0 \\ \int_a^b |f(x)|^p\, dx &= 0 \\ f(x)^p &= 0 \\ f(x) &= 0. \end{aligned}$$

The natural question to ask now, is what happens if $0 < p < 1$? The next proposition from [Roy88] tells us that we no longer have a normed space in this case, but we do have a metric space. For the sake of simplicity.

Let us denote

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

This suggests that $\|\cdot\|_p$ is a norm, which will be true once we make a slight modification to our space. We will not need any properties of it being a norm for the following proof, and we will explain that it is a norm in the section on L^p spaces.

Proposition 91 (Minkowski's Inequality for $0 < p < 1$)**.** *let* $f, g \in \mathcal{L}^p(X)$ *be non-negative functions and suppose* $0 < p < 1$*. Then*

$$\|f+g\|_p \geq \|f\|_p + \|g\|_p\,.$$

Proof. Since x^p is concave for $0 < p < 1$, let $t \in (0,1)$ and we have

$$(f+g)^p = \left(t\frac{f}{t} + (1-t)\frac{g}{(1-t)}\right)^p \geq t\frac{f^p}{t^p} + (1-t)\frac{g^p}{(1-t)^p}.$$

Integrating both sides gives

$$\|f+g\|_p^p \geq r\frac{\|f\|_p^p}{t^p} + (1-t)\frac{\|g\|_p^p}{(1-t)^p}.$$

Let

$$t = \frac{\|f\|_p}{\|f\|_p + \|g\|_p}, \quad 1-t = \frac{\|g\|_p}{\|f\|_p + \|g\|_p}.$$

Then

$$\begin{aligned}
\|f+g\|_p^p &\geq t\frac{\|f\|_p^p}{\dfrac{\|f\|_p^p}{\left(\|f\|_p + \|g\|_p\right)^p}} + (1-t)\frac{\|g\|_p^p}{\dfrac{\|g\|_p^p}{\left(\|f\|_p + \|g\|_p\right)^p}} \\
\|f+g\|_p^p &\geq t\left(\|f\|_p + \|g\|_p\right)^p + (1-t)\left(\|f\|_p + \|g\|_p\right)^p \\
\|f+g\|_p &\geq \ \|f\|_p + \|g\|_p .
\end{aligned}$$

□

Lemma 92. *A metric* d *induced by a norm on a normed space* X *satisfies*

(a) (translation invariance) $d(x+a, y+a) = d(x,y)$

(b) (scaling invariance) $d(\alpha x, \alpha y) = |\alpha|\, d(x,y)$

for all $x, y \in X$ *and every scalar* α.

Proof. We have

$$d(x+a, y+a) = \|x+a-(y+a)\| = \|x-y\| = d(x,y)$$

and

$$d(\alpha x, \alpha y) = \|\alpha x - \alpha y\| = |\alpha|\,\|x-y\| = |\alpha|\, d(x,y).$$

□

We can use this lemma to prove that some distances cannot arise from norms.

Example.

(a) Let X be a K-vector space and consider the distance defined by

$$d(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

Let $\alpha \neq 1$ and $\alpha \neq 0$ and $x, y \in X$ with $x \neq y$. Then by lemma 92, if this distance arose from a metric, we would have

$$d(\alpha x, \alpha y) = |\alpha|\, d(x,y)$$

but then $d(\alpha x, \alpha y) \neq 1$, which isn't true. Thus this metric cannot have come from a distance.

(b) Consider the sequence space s. Let $\alpha > 0$ and then we have

$$d(\alpha x, \alpha y) = \sum_{j=1}^{\infty} \frac{1}{2^j} \frac{|\alpha\xi_j - \alpha\eta_j|}{1 + |\alpha\xi_j - \alpha\eta_j|} = \sum_{j=1}^{\infty} \frac{1}{2^j} \frac{|\alpha| |\xi_j - \eta_j|}{1 + |\alpha| |\xi_j - \eta_j|} \neq |\alpha| d(x, y).$$

Proposition 93 ([Con90] Proposition 3.1.3)**.** *If* $(X, \|\cdot\|)$ *is a normed space, then*

(a) the function $X \times X \to X$ *defined by* $(x, y) \mapsto x + y$ *is continuous.*

(b) the function $K \times X \to X$ *defined by* $(\alpha, x) \mapsto \alpha x$ *is continuous.*

Proof.

(a) If $x_n \to x$ and $y_n \to y$, then $\|(x_n + y_n) - (x + y)\| = \|(x_n - x) + (y_n - y)\| \leq \|x_n - x\| + \|y_n - y\| \to 0$ as $n \to \infty$.

(b) If $x_n \to x$, then

$$\|\alpha x_n - \alpha x\| = \|\alpha(x_n - x)\| = |\alpha| \|x_n - x\| \to 0$$

as $n \to \infty$.

□

Definition 94. If $\|\cdot\|_1$ and $\|\cdot\|_2$ are two norms on a K-vector space X, they are said to be *equivalent norms* if they define the same topology on X.

Lemma 95 ([Con90] Lemma 3.1.4)**.** *If* p *and* q *are seminorms on a* K*-vector space* X*, then the following statements are equivalent.*

(a) $p(x) \leq q(x)$ *for all* x*.*

(b) $\{x \in X \mid q(x) < 1\} \subseteq \{x \in X \mid p(x) < 1\}$*.*

(c) $p(x) < 1$ *whenever* $q(x) < 1$*.*

(d) $\{x \in X \mid q(x) \leq 1\} \subseteq \{x \in X \mid p(x) \leq 1\}$*.*

(e) $p(x) \leq 1$ *whenever* $q(x) \leq 1$*.*

(f) $\{x \in X \mid q(x) < 1\} \subseteq \{x \in X \mid p(x) \leq 1\}$*.*

(g) $p(x) \leq 1$ *whenever* $q(x) < 1$*.*

Proposition 96 ([Con90] Proposition 3.1.5)**.** *If* $\|\cdot\|_1$ *and* $\|\cdot\|_2$ *are two norms on* X*, then these norms are equivalent if and only if there are positive constants* c *and* C *such that*

$$c \|x\|_1 \leq \|x\|_2 \leq C \|x\|_1$$

for all $x \in X$*.*

Proof. Suppose there exist constants c, C such that $c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1$ for all $x \in X$. Fix $x_0 \in X$ and let $\epsilon > 0$. Then

$$\{x \in X \mid \|x - x_0\|_1 < \epsilon/C\} \subseteq \{x \in X \mid \|x - x_0\|_2 < \epsilon\}, \{x \in X \mid \|x - x_0\|_2 < c\epsilon\} \subseteq \quad \{x \in X \mid \|x - x_0\|_1 < \epsilon\}.$$

This shows that the two topologies are the same.

Now suppose that the two norms are equivalent. Hence $\{x \in X \mid \|x\|_1 < 1\}$ is an open neighborhood of 0 in the topology defined by $\|\cdot\|_2$. Therefore there is an $r > 0$ such that $\{x \in X \mid \|x\|_2 < r\} \subseteq \{x \in X \mid \|x\|_1 < 1\}$. If $q(x) = r^{-1}\|x\|_2$ and $p(x) = \|x\|_1$, the preceding lemma implies $\|x\|_1 \leq r^{-1}\|x\|_2$ or $c\|x\|_1 \leq \|x\|_2$, where $c = r$.

For the other inequality, $\{x \in X \mid \|x\|_2 < 1\}$ is an open neighborhood of 0 in the topology defined by $\|\cdot\|_1$. Therefore, there is an $s > 0$ such that

$$\{x \in X \mid \|x\|_1 < s\} \subseteq \{x \in X \mid \|x\|_2 < 1\}.$$

If $q(x).s^{-1}\|x\|_1$ and $p(x) = \|x\|_2$, the preceding lemma implies $\|x\|_2 \leq s^{-1}\|x\|_1$. Taking $\frac{1}{C} = s$, we get the inequality. ☐

Example. Take $X = \mathbb{R}$ and we have the two norms ℓ^1 and ℓ^2, taking $c = \frac{1}{\sqrt{2}}$ and $C = 1$, we have

$$\frac{1}{\sqrt{2}}(|x_1| + |x_2|) \leq \sqrt{x_1^2 + x_2^2} \leq |x_1| + |x_2|.$$

We get the $C = 1$ from the triangle inequality, and the $\frac{1}{\sqrt{2}}$ can be obtained from the Cauchy-Schwartz inequality:

$$|\vec{x}| = \sum_{i=1}^{2} |x_i| \cdot 1 \leq \left(\sum_{i=1}^{2} |x_i|^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{2} 1^2\right)^{\frac{1}{2}} = \sqrt{x_1^2 + x_2^2}\sqrt{2}.$$

Replacing 2 by p in the above generalizes the proof to showing that for any ℓ^p is equivalent to ℓ^1 for any $p < \infty$, the only difference is that $c = \frac{1}{\sqrt{n}}$.

Our next goal is to show that all norms on a finite dimensional K-vector space are equivalent.

Lemma 97. *Let $\{x_1, \dots, x_n\}$ be a linearly independent set of vectors in a normed space X. Then there is a number $c > 0$ such that for every choice of scalars $\alpha_1, \dots, \alpha_n$, we have*

$$\|\alpha_1 x_1 + \cdots + \alpha_n x_n\| \geq c(|\alpha_1| + \cdots + |\alpha_n|).$$

Theorem 98 ([Con90] Theorem 3.3.1)**.** *On a finite dimensional vector space X, any norm $\|\cdot\|$ is equivalent to any other norm $\|\cdot\|_0$.*

Proof. Let $\dim X = n$ and $\{e_1, \dots, e_n\}$ be any basis for X. Ten every $x \in X$ has a unique representation

$$x = \alpha_1 e_1 + \cdots + \alpha_n e_n.$$

By lemma there is a positive constant c such that

$$\|x\| \geq c(|\alpha_1| + \cdots + |\alpha_n|).$$

On the other hand the triangle inequality gives

$$\|x\|_0 \leq \sum_{j=0}^{n} |\alpha_j| \, \|e_j\|_0 \leq k \sum_{j=1}^{n} |\alpha_j| \, ,$$

where $k = \max\left\{\|e_j\|_0 \mid j = 1, \ldots, n\right\}$. Putting these together we get that $\frac{c}{k}\|x\|_0 \leq \|x\|$. We can get the other direction by switching $\|\cdot\|$ and $\|\cdot\|_0$ in the above proof. □

Theorem 99 (Parallelogram Law). *If* $(X, \|\cdot\|)$ *is a normed space and* $f, g \in X$*, then*

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|)^2.$$

Remark 100. This theorem says that the sums of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the four sides, as can be seen in figure.

Another important inequality of this type is the following:

Theorem 101 (Ptolemy's Law). *Let* $(X, \|\cdot\|)$ *be a normed space. Then* $\|x\|$ *induces an inner product is and only if*

$$\|x-y\| \, \|z\| + \|y-z\| \, \|x\| \geq \|x-z\| \, \|y\| \, .$$

Remark 102. We finish this section with a quick remark about the difference between metric spaces and normed spaces. First, to have a norm on a space X, we need that X is a vector space. This means that we can add and subtract elements and multiply by scalars. We can move elements around. In order to have a metric on a space X, we don't need to be able t do this. We can put a metric on a much more general space, one with much less structure. As an example, for any space X we can put the trivial metric on it:

$$d(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

Our space does not need any structure to put a metric on it, since all that a metric does is measure the distance between two elements of our set. On the other hand, a norm measures the *size* of elements of our space, and thus can induce a distance since if we know the size of two objects we can find the size of their difference, which is their distance.

Chapter 10

Quotient Spaces

Our goal in this section is to build the necessary background needed to defined the L^p spaces. We start with basic definitions of relations and equivalence relations, and then move on to quotient spaces and give many examples. More on relations can be found in [Vel06].

Definition 103. Suppose A and B are sets. Then the *Cartesian product* of A and B, denoted $A \times B$ is the set of all ordered pairs whose first coordinate is an element from A and whose second coordinate is an element from B. We write this as

$$A \times B = \{(a, b) \mid a \in A, b \in B\}.$$

Example.

(a) If $A = \{1, 2, 3\}$ and $B = \{4, 5\}$, then $A \times B = \{(1, 4), (1, 5), (2, 4), (2, 5), (3, 4), (3, 5)\}$.

(b) If $A = B = \mathbb{R}$, then $A \times B = \mathbb{R}^2$, the real plane.

(c)

Definition 104. Suppose A and B are sets. Then a set $R \subseteq A \times B$ is called a *relation* from A to B.

Example.

(a) If $A = \{1, 2, 3\}, B = \{4, 5\}$, then $R = \{(1, 4), (3, 5)\}$ is a relation from A to B.

(b) Let $A = B = \mathbb{R}$, then $G = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x > y\}$ is a relation from $\mathbb{R}$ to $\mathbb{R}$.

We can actually define functions in terms of relations.

Definition 105. Suppose F is a relation from A to B. Then F is a *function* from A to B if for every $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in F$.

Example.

(a) Let $A = \{1, 2, 3\}, B = \{4, 5, 6\}$ and $F = \{(1, 5), (2, 4), (3, 5)\}$, then F is a function from A to B.

(b) Let $A = B = \mathbb{R}$, and let $F = \left\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x^2 = y\right\}$, then this is a relation from $\mathbb{R}$ to $\mathbb{R}$.

Definition 106. Let R be a relation from A to A such that

(a) (reflexive) For all $x \in A$, $(x, x) \in R$.

(b) (symmetric) for all $(x, y) \in R$, $(y, x) \in R$.

(c) (transitive) if $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$,

then we call such a relation an *equivalence relation*.

Definition 107. Suppose R is an equivalence relation on a set A, and $x \in A$. Then the *equivalence class of* x *with respect to* R is the set

$$[x]_R = \{y \in A \mid (y, x) \in R\}.$$

Remark 108. From now on we will often denote an equivalence relation by $\sim$, and say $x \sim y$ for $(x, y) \in R$.

Definition 109. The set of all equivalence classes of elements of A is called the *quotient of* A *with* $\sim$, or A *modulo* $\sim$, and is denoted $A/\sim$. Thus,

$$A/\sim = \{[x]_\sim \mid x \in A\}.$$

Example.

(a) Let X be a K-vector space, and let Y be a nonempty subspace. Define $x \sim y$ is $x - y \in Y$. Then $\sim$ is an equivalence relation. Indeed, $x \sim x$ since $x - x = 0 \in Y$. It is always the case that 0 is in Y, since Y is itself a nonempty vector space, so there is a $y \in Y$, then $(-1) \cdot y = -y \in Y$ as well since Y is linear, but then $y + (-y) = 0 \in Y$. Next, if $x \sim y$, then $x - y \in Y$, so $(-1)(x - y) = y - x \in Y$, thus $y \sim x$. Finally, suppose that $x \sim y$ and $y \sim x$, then $x - y \in Y$ and $y - z \in Y$, so $(x - y) + (y - z) = x - z \in Y$. Thus $\sim$ is an equivalence relation. Then $X/Y = \{[x] \mid x \in X\}$ where $[x] = \{y \in Y \mid x \sim y\}$. We can put the structure of a K-vector space on X/Y in the following way. Define $[x] + [y]$ by choosing $x_1 \in [x]$ and $y_1 \in [y]$ and letting $[x] + [y] = [x_1 + y_1]$. We need to check that this is well defined. This means that if we make a different choice for x_1 and y_1, we still end with the same equivalence classes. Suppose $x_2 \neq x_1$ and $x_2 \in [x]$, and $y_2 \in [y]$ with $y_1 \neq y_2$. Take $z \in [x_1 + y_1]$, we will show that $z \in [x_2 + y_1]$. Starting with $z - (x_1 + y_1) \in Y$ we have

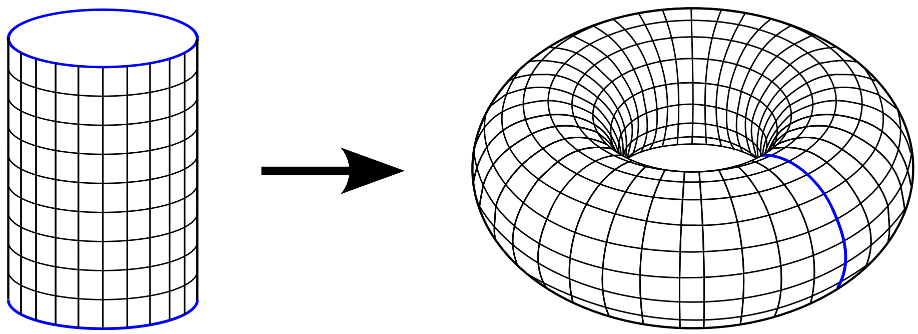
$$\begin{aligned} z - (x_1 + y_1) &= z - (x_2 + y_2) + (x_2 + y_2) - (x_1 + y_1) \in Y \\ &= z - (x_2 + y_2) + (x_2 + y_2) - (x_1 + y_1) - (x_2 - x_1) - (y_2 - y_1) \in Y \\ & z - (x_2 + y_2) \in Y. \end{aligned}$$

Thus $z \in [x_2 + y_2]$. Since this argument is symmetric, we get $[x_1 + y_1] = [x_2 + y_2]$. Similarly, we define $\alpha \cdot [x]$ to be $[\alpha x_1]$ for some $x_1 \in [x]$. This can be shown to be well-defined in a similar manner. Thus X/Y is a K-vector space. Furthermore, if X has a norm $\|\cdot\|$, and Y is closed, then we can put a norm on X/Y by letting

$$\|[x]\| = \inf\{\|z\| \mid z \in [x]\}.$$

(b) The above construction works for groups as well. Given an abelian group G, and a subgroup H, the quotient is $G/H = \{x + H \mid x \in G\}$ where $x + H = [x]$.

(c) Quotients appear in topology as well. Take a unit square, $[0, 1] \times [0, 1]$. We can identify parts of the square to get a *quotient space*. We say $(0, x) \sim (1, x)$, in this way we are identifying the two sides of a square, and gluing them together to get a new space, in this case it will be a hollow cylinder. We can again identify the top circle with the lower circle and glue those together to get a *torus*:



Chapter 11

Complete Spaces

In this section we formalize completion. This is the process that takes us from $\mathbb{Q}$ to $\mathbb{R}$, but this is applicable in far more general metric spaces.

Definition 110. A sequence $\{x_n\}$ in a metric space X is said to be a *Cauchy sequence* if for every $\epsilon > 0$ there is an integer N such that $d(x_n, x_m) < \epsilon$ if $n, m \geq N$.

Example. The sequence $\left\{\frac{1}{n}\right\}_{n=1}^{\infty}$ in $\mathbb{R}$ is a Cauchy sequence. Indeed, let $\epsilon > 0$. Set $N > \frac{2}{\epsilon}$. Then for $m, n \geq N$, we have that $\frac{1}{m} < \frac{\epsilon}{2}$ and $\frac{1}{n} < \frac{\epsilon}{2}$. Then

$$\left|\frac{1}{n} - \frac{1}{m}\right| \leq \left|\frac{1}{n}\right| + \left|\frac{1}{m}\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus the sequence is Cauchy, and we know that is converges to 0.

Example.

(a) Not all Cauchy sequences converge. For example, if we take our metric space to be $X = (0, 1]$, then the sequence $\left\{\frac{1}{N}\right\}_{n=1}^{\infty}$ no longer converges, since $0 \notin X$.

(b) Put in the example of continuous functions on $[0, 1]$ that converge to a discontinuous function. This will be a Cauchy sequence that does not converge in $C([0, 1], \mathbb{R})$.

Definition 111. Let E be a subset of a metric space X, and let

$$S = \{d(x, y) \mid x, y \in E\}.$$

We call $\sup S$, the *diameter*, and it will be denoted as $\operatorname{diam}(E_N)$.

Remark 112. Note that if $\{x_n\}$ is a sequence in X and $E_N = \{x_n \mid n \geq N\}$, then $\{x_n\}$ is a Cauchy sequence if and only if

$$\lim_{N\to\infty} \operatorname{diam}(E_N) = 0.$$

We now prove two theorems that will help tell us when Cauchy sequences converge.

Theorem 113 ([Rud76] Theorem 3.10)**.**

(a) If $\overline{E}$ is the closure of a set E in a metric space X, then

$$\operatorname{diam}(\overline{E}) = \operatorname{diam}(E).$$

(b) If K_n is a sequence of compact sets in X such that $K_n \supset K_{n+1}$ and if

$$\lim_{n\to\infty} \operatorname{diam}(K_n) = 0,$$

then $\bigcap_{n=1}^{\infty} K_n$ consists of exactly one point.

Proof.

(a) Since $E \subseteq \overline{E}$, clearly $\operatorname{diam}(E) \leq \operatorname{diam}(\overline{E})$. Suppose $\epsilon > 0$, and let $x, y \in \overline{E}$. There exist points $x', y' \in \overline{E}$ such that $d(x, x') < \epsilon$, $d(y, y') < \epsilon$. Hence

$$d(x,y) \leq d(x,x') + d(x',y') + d(y',y) < 2\epsilon + d(x',y') \leq 2\epsilon + \operatorname{diam}(E).$$

It follows that $\operatorname{diam}(\overline{E}) \leq 2\epsilon + \operatorname{diam}(E)$.

(b) Put $K = \bigcap_{n=1}^{\infty} K_n$. Since the K_n are compact, K is nonempty. If K contains more than one point, then $\operatorname{diam}(K) > 0$, but for each n, $K_n \supset K$, so $\operatorname{diam}(K_n) \geq \operatorname{diam}(K)$. This is a contradiction to the fact that $\operatorname{diam}(K_n) \to 0$.

□

Theorem 114 ([Rud76] Theorem 3.11)**.**

(a) In any metric space X, every convergent sequence is a Cauchy sequence.

(b) If X is a compact metric space and if $\{x_n\}$ is a Cauchy sequence in X, then $\{x_n\}$ converges to some point in X.

(c) In $\mathbb{R}^n$, every Cauchy sequence converges.

Proof.

(a) Suppose $x_n \to x$, and let $\epsilon > 0$. Then there is an integer $N > 0$ such that $d(x, x_n) < \epsilon/2$ for all $n \geq N$. Hence

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon$$

whenever $n, m \geq N$. Thus $\{x_n\}$ is a Cauchy sequence.

(b) Let $\{x_n\}$ be a Cauchy sequence in a compact space X. For $N = 1, 2, \ldots$, let E_N be the set consisting of $x_N, x_{N+1}, \ldots$. Then by theorem 113

$$\lim_{N\to\infty} \operatorname{diam}(\overline{E_N}) = 0.$$

Being a closed subset of the compact space X, each $\overline{E_N}$ is compact, and $E_N \supset E_{N+1}$, so $\overline{E_N} \supset \overline{E_{N+1}}$. Theorem 113, shows that there is a unique $x \in X$ which lies in every $\overline{E_N}$. Let $\epsilon > 0$. There is an integer N_0 such that $\operatorname{diam}(\overline{E_N}) < \epsilon$ if $N \geq N_0$. Since $x \in \overline{E_N}$, it follows that $d(x,y) < \epsilon$ for every $y \in \overline{E_N}$, hence for every $y \in E_N$. In other words $d(x, x_n) < \epsilon$ if $n \geq N_0$. This is saying that $x_n \to x$, and thus it converges.

(c) Let $\{x_n\}$ be a Cauchy sequence in $\mathbb{R}^n$. Define E_n as above. For some $N > 0$, $\operatorname{diam} E_N < 1$. The range of $\{x_n\}$ is the union of E_N and the finite set $\{x_1, \dots, x_{N-1}\}$. Hence $\{x_n\}$ is bounded. Since every bounded subset of $\mathbb{R}^n$ has a compact closure in $\mathbb{R}^n$, this follows from (2).

□

Definition 115. A metric space in which every Cauchy sequence converges is called *complete*.

Definition 116. Let (X, d) and $(\tilde{X}, \tilde{d})$ be metric spaces. Then

(a) A mapping T of X into $\tilde{X}$ is said to be an *isometry* if T preserves distances, that is, for all $x, y \in X$,

$$\tilde{d}(T(x), T(y)) = d(x, y)$$

(b) The space X is said to be isometric with the space $\tilde{X}$ if there exists a bijective isometry of X onto $\tilde{X}$.

Example.

(a) Take $X = \tilde{X} = \mathbb{R}$ with the metric given by the usual absolute value. Then for any $k \in \mathbb{R}$, the map $T\colon \mathbb{R} \to \mathbb{R}$ given by $T(x) = x + k$ is an isometry. Indeed, let $x, y \in \mathbb{R}$, then

$$d(T(x), T(y)) = |T(x) - T(y)| = |x + k - (y + k)| = |x - y| = d(x, y).$$

(b) Similarly, in $\mathbb{R}^n$ for any vector $\vec{x} \in \mathbb{R}^n$, the map defined by $T(\vec{y}) = \vec{y} + \vec{x}$ is an isometry.

Theorem 117 ([Kre78] Theorem 1.6-2)**.** *For a metric space $X = (X, d)$ there exists a complete metric space $(\widehat{X}, \widehat{d})$ which has a subspace W that is isometric with X and is dense in $\widehat{X}$. This space $\widehat{X}$ is unique except for isometries, that is, if $\widetilde{X}$ is any complete metric space having a dense subspace $\widetilde{W}$ isometric with X, then $\widetilde{X}$ and $\widehat{X}$ are isometric.*

Proof. We describe the construction of $\widehat{X}$ here and the rest of the details can be found in [Kre78], Theorem 1.6-2. Let (x_n) and (x_n') be Cauchy sequences in X. Define $(x_n) \sim (x_n')$ if

$$\lim_{n\to\infty} d(x_n, x_n') = 0.$$

Let $\widehat{X}$ be the set of all equivalence classes $\widehat{x}, \widehat{y}, \dots$ of Cauchy sequences. We now set

$$\widehat{d} = \lim_{n\to\infty} d(x_n, y_n)$$

where $(x_n) \in \widehat{x}$ and $(y_n) \in \widehat{y}$. We quickly show that this limit exists. We have

$$d(x_n, y_n) \le d(x_n, x_m) + d(x_m, y_m) + d(y_m, y_n)$$

and thus

$$d(x_n, y_n) - d(x_m, y_m) \le d(x_n, x_m) + d(y_m, y_n)$$

and we can do the same with m and n swapped. Together this gives

$$|d(x_n, y_n) - d(x_m, y_m)| \le d(x_n, x_m) + d(y_m, y_n),$$

and since (x_n) and (y_n) are Cauchy, we can make the right hand side as small as we want, and thus the limit exists since we have already proven that $\mathbb{R}$ is complete.

Next, we show that the limit is independent of the choice of representatives. Suppose $(x_n) \sim (x'_n)$ and $(y_n) \sim (y'_n)$. Then

$$\left|d(x)n, y_n) - d(x'_n - y'_n)\right| \leq d(x_n, x'_n) + d(y_n, y'_n) \to 0$$

as $n \to \infty$, which implies that

$$\lim_{n\to\infty} d(x_n, y_n) = \lim n \to \infty d(x'_n, y'_n).$$

Lastly, we verify that $\widehat{d}$ is indeed a metric. Clearly it is always non-negative, and $d(\widehat{x}, \widehat{x}) = 0$. If $\widehat{d}(\widehat{x}, \widehat{y}) = 0$, then we have that $(x_n) \sim (y_n)$, and so $\widehat{x} = \widehat{y}$. Finally, the triangle inequality follows from

$$(x_n, y_n) \leq d(x_n, z_n) + d(z_n, y_n).$$

□

Example.

(a) If $X = \mathbb{Q}$, then the completion of $\mathbb{Q}$ with respect to the absolute value is $\mathbb{R}$. This means that we should really think of $\mathbb{R}$ as a set of equivalence classes of Cauchy sequences.

(b) Let $X = \mathbb{Q}$. We will complete $\mathbb{Q}$ again, but this time we will use a different metric. Fix a prime number p. For any integer n, let $\nu_p(n)$ be the highest power of p that divides n. So if $n = 12$ and $p = 2$, then $\nu_2(12) = 2$. Then extend ν to $\mathbb{Q}$ by setting

$$\nu_p\left(\frac{a}{b}\right) = \nu(a) - \nu(b).$$

For example

$$\nu_2\left(\frac{1}{12}\right) = \nu(1) - \nu(12) = 0 - 2 = -2.$$

Now, we define the p-*adic absolute value* to be

$$\left|\frac{a}{b}\right|_p = p^{-\nu(a/b)}.$$

So

$$\left|\frac{1}{12}\right|_2 = 2^{-2} = \frac{1}{4}.$$

Now we are considering a number to be small if it is divisible by p many times, whereas it is small if it is not. While this choice of absolute value seems arbitrary, Ostrowski's theorem says that the only absolute values on $\mathbb{Q}$ are the usual one $|\cdot|$, and the p-adic ones, $|\cdot|_p$. Using the p-adic absolute values give rise to all sorts of strange geometric properties such as all triangles are isosceles, and every point of a ball is a center of the ball. Now, if we complete $\mathbb{Q}$ with respect to a p-adic absolute value, i.e., do the same procedure as in the proof of theorem 117, but replace every absolute value with the p-adic absolute value, we will get the field of p-adic number, $\mathbb{Q}_p$.

Now we give some examples of complete spaces

Example.

(a) Let X be the set of all bounded sequences of real numbers. So if $x \in X$, then $x = (\xi_i)_{i=1}^{\infty}$, and for each $x \in X$, there is some $c_x \in \mathbb{R}$ such that $|\xi_i| \leq c_x$. We denote this space by l^∞. A norm on this space is

$$\|(\xi_i)\| = \sup\{|\xi_i| \mid i \geq 1\}$$

and thus the induced distance is

$$d(x, y) = \sup\{|\xi_i - \eta_i| \mid i \geq 1\}$$

where $x = (\xi_i), y = (\eta_i)$. Now we show that this space is complete. Let (x_m) be a Cauchy sequence in l^∞, where $x_m = (\xi_i^m)$. Given $\epsilon > 0$, there is an $N > 0$ such that for all $m, n \geq N$,

$$d(x_m, x_n) = \sup\{|\xi_i^m - \xi_i^n| \mid i \geq 1\} < \epsilon.$$

So, for every $i \geq 1$ and $n, m \geq N$,

$$|\xi_i^m - \xi_i^n| < \epsilon.$$

Since $\mathbb{R}$ is complete, the sequence $(\xi_i^j)_{j=1\infty}$ is a Cauchy sequence and thus it converges to some ξ_i. We define $x = (\xi_i)_i = 1^\infty$. Now, we have

$$\left|\xi_j^m - \xi_j\right| < \epsilon$$

for $m \geq N$. Since $x_m = (\xi_i^m) \in l^\infty$, there is a real number k_m such that $\left|\xi_j^m\right| \leq k_m$ for all j. Hence,

$$|\xi| \leq |\xi_i - \xi_i^m| + |\xi_i^m| \leq \epsilon + k_m.$$

This inequality holds for every j, and the right hand side does not involve j. Hence (ξ_i) is bounded, and this it is in l^∞. Also, we have

$$d(x_m, x) = \sup\{|\xi_i^m - \xi_i| \mid i \geq 1\} \leq \epsilon,$$

which shows that $x_m \to x$, and thus l^∞ is complete.

(b) Now consider the space of continuous functions $f\colon [a, b] \to \mathbb{R}$, denote $C(a, b], \mathbb{R})$. The norm we put on this space is

$$\|f\|_\infty = \max\{|f(x)| \mid x \in [a, b]\},$$

and thus the induced distance is

$$d(f, g) = \max\{|f(x) - g(x)| \mid x \in [a, b]\}.$$

We will prove that this space is complete. Let (f_m) be a Cauchy sequence in $C([a, b], \mathbb{R})$. Then given $\epsilon > 0$ there exists an $N > 0$ such that for all $m, n > N$ we have

$$d(f_m, f_n) = \max\{|f_m(x) - f_n(x)| \mid x \in [a, b]\} < \epsilon.$$

Hence for any fixed $x_0 \in [a, b]$,

$$|f_m(x_0) - f_n(x_0)| < \epsilon.$$

This shows that $(f_m(x_0))_{m=1}^{\infty}$ is a Cauchy sequence of real numbers, which converges since $\mathbb{R}$ is complete. Say the sequence converges to $f(x_0)$ as $m \to \infty$. So for each $x \in [a, b]$ we can associate a unique real number $f(x)$. This defines a function on $[a, b]$. To see that $f \in C([a, b], \mathbb{R})$, we have

$$\max\{|f_m(x) - f(x)| \mid x \in [a, b]\} < \epsilon.$$

Hence for each $x \in [a, b]$, we have

$$|f_m(x) - f(x)| < \epsilon.$$

This shows that $(f_m(x))$ converges to $f(x)$ uniformly on $[a, b]$. Since the f_m are continuous on $[a, b]$ and the convergence is uniform, f is continuous on $[a, b]$ as well. Thus $f \in C([a, b], \mathbb{R})$, and $f_m \to f$, and the space is complete.

(c) To show that completeness is dependent on the metric, consider the space $C([0, 1], \mathbb{R})$ under the metric

$$d(f, g) = \int_0^1 |f(x) - g(x)|\, dx.$$

We will show that this space is not complete. Consider the sequence of functions defined by

$$f_m(t) = \begin{cases} 0 & \text{if } x \in [0, \frac{1}{2}] \\ mx - \frac{m}{2} & \text{if } x \in (\frac{1}{2}, \frac{1}{2} + \frac{1}{m}] \\ 1 & \text{if } x \in (\frac{1}{2} + \frac{1}{m}, 1] \end{cases}$$

This sequence of functions defines a Cauchy sequence, since the only place where they are not equal is on the interval $[\frac{1}{2}, \frac{1}{2} + \frac{1}{n}]$ which goes to 0. We will show that this Cauchy sequence does not converge. For every $f \in C([a, b], \mathbb{R})$ we have

$$d(f_m, f) = \int_0^1 |f_m(x) - f(x)|\, dx = \int_0^1 |f(x)|\, dx + \int_{1/2}^{1/2+1/m} |f_m(x) - f(x)|\, dx + \int_{1/2+1/n}^1 |1 - f(x)|\, dx$$

Since the integrands are non-negative, so are the integrals. So if $d(f_m, f) \to 0$, this would mean that each integral approaches 0, and since f is continuous, we should have

$$f(x) = 0, \text{ for } x \in [0, \frac{1}{2}), \quad f(X) = 1 \text{ if } x \in (\frac{1}{2}, 1]$$

but this is not possible for a continuous function. Thus the sequence (f_m) does not converge and the space is not complete. This example illustrates that a space X can be complete with respect to one metric, but not complete with respect to another.

Now we prove a result that gives us another criterion for proving that a space is complete.

Definition 118. A series $\{f_n\}$ in a normed linear space $(X, \|\cdot\|)$ is said to be *summable* to a sum s if $s \in X$ and the sequence of partial sums of the series converges to s, that is

$$\left\| s - \sum_{i=1}^{n} f_i \right\| \to 0.$$

If this is the case, we write

$$s = \sum_{i=1}^{\infty} f_i.$$

The series is said to be *absolutely summable* if $\sum_{i=1}^{\infty} \|f_i\| < \infty$.

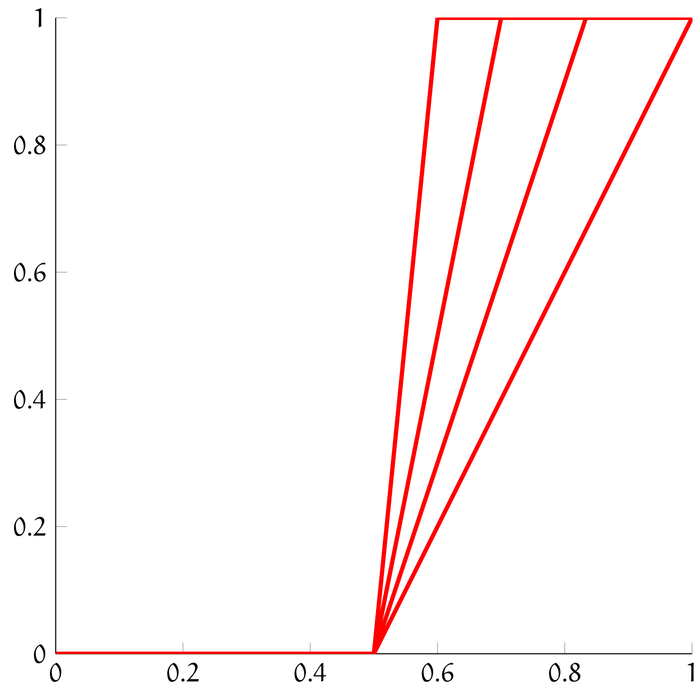


Figure 11.1: The functions f_2, f_3, f_5, and f_{10}.

Proposition 119 ([Roy88]). *A normed linear space* $(X, \|\cdot\|)$ *is complete if and only if every absolutely summable series is summable.*

Proof. Suppose that X is complete and let $\{f_n\}$ be an absolutely summable series of elements of X. Since $\sum \|f_i\| = M < \infty$, there is, for each $\epsilon > 0$, an N such that $\sum_{i=N}^{\infty} \|f_i\| < \epsilon$. Let $s_n = \sum_{i=1}^{n} f_i$ be the partial sum of the series. Then for $n \geq m \geq N$ we have

$$\|s_n - s_m\| = \left\| \sum_{i=m}^{n} f_i \right\| \leq \sum_{i=1}^{\infty} \|f_i\| \leq \sum_{i=N}^{\infty} < \epsilon.$$

Hence the sequence $\{s_n\}$ of partial sums is a Cauchy sequence and must converge to an element s in X, since X is complete.

For the other direction, let $\{f_n\}$ be a Cauchy sequence in X. For each integer k there is an integer n_k such that $\|f_n - f_m\| < 2^{-k}$ for all n and m greater than n_k, and we may choose the n_k's so that $n_{k+1} > n_k$. Then $\{f_{n_k}\}_{k=1}^{\infty}$ is a subsequence of $\{f_n\}$, and if we set $g_1 = f_{n_1}$ and $g_k = f_{n_k} - f_{n_{k-1}}$ for $k > 1$ we get a series $\{g_k\}$ whose kth partial sum is f_{n_k}. We have $\|g_k\| < 2^{-k+1}$ if $k > 1$, Thus

$$\sum \|g_k\| \leq \|g_1\| + \sum 2^{-k+1} = \|g_1\| + 1.$$

Hence the series $\{g_k\}$ is absolutely summable, and so by our hypothesis there is an element f in X to which the partial sums of the series converge. Thus the subsequence $\{f_{n_k}\}$ converges to f.

Now we show that $f = \lim_{n\to\infty} f_n$. Since $\{f_n\}$ is a Cauchy sequence, given $\epsilon > 0$, there is an N such that $\|f_n - f_m\| < \epsilon/2$ for all n and m large than N. Since $f_{n_k} \to f$, there is a K such that for all $k \geq K$ we have

$\|f_{n_k} - f\| < \epsilon/2$. Let us take k so large that $k > K$ and $n_k > N$. Then

$$\|f_n - f\| \leq \|f - f_{n_k}\| + \|f_{n_k} - f\| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus for all $n > N$ we have $\|f_n - f\| < \epsilon$ and so $f_n \to f$. □

Chapter 12

Banach Spaces

We have already seen many examples of Banach spaces, and now we formally define them.

Definition 120. A Banach space is a normed space $(X, \|\cdot\|)$ that is complete with respect to the metric defined by the norm:

$$d(x, y) = \|x - y\|.$$

Any example of a complete normed space we have seen in the past few sections is a Banach space, $\mathbb{R}^n, \mathbb{C}$, and l^∞. We provide a few more examples.

Example.

(a) Let $n \geq 1$ and let $C^{(n)}[0, 1]$ be the collection of functions $f: [0, 1] \to \mathbb{R}$ such that f has n continuous derivatives. Define

$$\|f\| = \sup\left\{\sup\left\{\left|f^{(k)}(x)\right| \mid 0 \leq x \leq 1\right\} \mid 0 \leq k \leq n\right\}.$$

(b) We have shown previously that $(C([a, b], \mathbb{R}), \|\cdot\|_\infty)$ is a complete normed space, and hence it is a Banach space.

(c) The space $(C([a, b], \mathbb{R}), \|\cdot\|_p)$ for $1 \leq p < \infty$ is not complete, and thus is not a Banach space. Recall that we can create a sequence of continuous functions that converge to a non-continuous function which is not in $C([a, b], \mathbb{R})$.

(d) The set of rational numbers, $\mathbb{Q}$ is not complete, but it is a normed space by setting $\|q\|_{\mathbb{Q}} = |q|$. We can see that it is not complete by considering the sequence defined as $q_1 = 1$, and $q_{n+1} = \frac{1}{2}\left(q_n + \frac{2}{q_n}\right)$ which will converge to $\sqrt{2}$. But, as is well known in a real analysis course, the completion of $\mathbb{Q}$ with respect to this norm is $\mathbb{R}$.

Definition 121. Let $(X, \|\cdot\|)$ be a normed space and let Y be a closed subspace. Let $Q: X \to X/Y$ be the natural map that sends $x \mapsto [x]$. Recall that we can define a norm on X/Y as

$$\|[x]\| = \inf\{\|z\| \mid z \in [x]\}.$$

Chapter 13

Hilbert Spaces

Hilbert Spaces are one of the main objects of study in functional analysis. They are an abstraction of the familiar Euclidean spaces. Throughout this section we denote by $\mathbb{K}$, either $\mathbb{R}$ or $\mathbb{C}$, since most results will hold in either case.

Definition 122. If X is a $\mathbb{K}$-vector space, a *inner product* on X is a function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$ such that for all $\alpha, \beta \in \mathbb{K}$ and all $x, y, z \in X$, the following are satisfied:

(a) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$,

(b) $\langle x, \alpha y + \beta z \rangle = \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle$,

(c) $\langle x, x \rangle \geq 0$ and if $\langle x, x \rangle = 0$, then $x = 0$,

(d) $\langle x, y \rangle = \overline{\langle y, x \rangle}$,

where $\overline{\alpha}$ is the complex conjugate of α, i.e., if $\alpha = a + bi$, then $\overline{\alpha} a - bi$.

Example.

(a) Let X be the collection of all sequences $\{\alpha_n\}$ of scalars $\alpha_n \in \mathbb{K}$ such that $\alpha_n = 0$ for all but a finite number of values of n. We can define addition and scalar multiplication as

$$\{\alpha_n\} + \{\beta_n\} = \{\alpha_n + \beta_n\}$$
$$\alpha\{\alpha_n\} = \{\alpha\alpha_n\}.$$

Then X is a $\mathbb{K}$-vector space. Each of the following is an inner product on X:

$$\langle \{\alpha_n\}, \{\beta_n\} \rangle = \sum_{n=1}^{\infty} \alpha_n \overline{\beta_n},$$
$$\langle \{\alpha_n\}, \{\beta_n\} \rangle = \sum_{n=1}^{\infty} \frac{1}{n} \alpha_n \overline{\beta_n},$$
$$\langle \{\alpha_n\}, \{\beta_n\} \rangle = \sum_{n=1}^{\infty} n^5 \alpha_n \overline{\beta_n}.$$

(b) Let $f, g \in \mathcal{L}^2(X)$, then Hölder's inequality says that $f\overline{g} \in \mathcal{L}^1(X)$. Setting

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

makes $\mathcal{L}^2(X)$ into an inner product space.

Definition 123. A *Hilbert space* is a vector space X over $\mathbb{K}$ together with an inner product $\langle \cdot, \cdot \rangle$ such that relative to the metric $d(x, y) = \|x - y\|$ induced by the norm, X is a complete metric space.

Example. Many of the examples we have already seen are Hilbert spaces.

(a) $\mathbb{R}^n$ with the dot product is a Hilbert space.

(b) $\mathbb{C}^n$ is a Hilbert space with the inner product defined by

$$\langle x, y \rangle = \xi_i\overline{\eta_i} + \cdots + \xi_n\overline{\eta_n}$$

where $x = (\xi_1, \ldots, \xi_n)$ and $y = (\eta_1, \ldots, \eta_n)$.

(c) We will soon define the L^p spaces where the space L^2 will be a very important Hilbert space. For all other $p \geq 1$ L^p will be a normed space.

(d) The space of sequences in $\mathbb{C}$, l^2 with inner product defined by

$$\langle x, y \rangle = \sum_{j=1}^{\infty} \xi_j\overline{\eta_j}$$

is a Hilbert space, but for every other $p \neq 2$, l^p is not a Hilbert space.

The following is a fundamental inequality in inner product spaces.

Theorem 124 (Cauchy-Bunyakowsky-Schwarz Inequality, [Con90] Theorem 1.1.4)**.** *If $\langle \cdot, \cdot \rangle$ is an inner product on X, then*

$$|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle$$

for all $x, y \in X$. Moreover, equality occurs if and only if there are scalars α, β both not 0, such that

$$\langle \beta x + \alpha y, \beta x + \alpha y \rangle = 0.$$

Proof. If $\alpha \in \mathbb{K}$ and $x, y \in X$, then

$$\begin{aligned} 0 \leq\ & \langle x - \alpha y, x - \alpha y \rangle \\ =\ & \langle x, x \rangle - \alpha \langle y, x \rangle - \overline{\alpha} \langle x, y \rangle + |\alpha|^2 \langle y, y \rangle . \end{aligned}$$

Suppose $\langle y, x \rangle = be^{i\theta}$ for some $b \geq 0, b \in \mathbb{R}$, and let $\alpha = te^{-i\theta}$ for $t \in \mathbb{R}$. The above inequality becomes

$$\begin{aligned} 0 \leq\ & \langle x, x \rangle - te^{-i\theta}be^{i\theta} - te^{i\theta}be^{-i\theta} + t^2 \langle y, x \rangle \\ =\ & \langle x, x \rangle - 2bt + t^2 \langle y, y \rangle \\ =\ & c - 2bt + at^2 = q(t) \end{aligned}$$

where $c = \langle x, x \rangle$ and $a = \langle y, y \rangle$. Thus $q(t)$ is a quadratic polynomial in t, and $q(t) \geq 0$ for all t. This implies that $q(t)$ has at most one real root. Using the discriminant, which we see is not positive, gives

$$0 \geq b^2 - ac = |\langle x, y \rangle|^2 - \langle x, x \rangle \langle y, y \rangle$$

proving the inequality.

Now, suppose that there exist $\alpha, \beta \in \mathbb{K}$ such that $\langle \alpha x + \beta y, \alpha x + \beta y \rangle = 0$. By the properties of an inner product we get

$$\alpha x + \beta y = 0 \implies y = -\frac{\beta}{\alpha} x.$$

Let $\lambda = -\frac{\beta}{\alpha}$, so that $y = \lambda x$. We see that the inequality is an equality if and only if the last line is an equality. We have

$$\begin{aligned} 0 &\geq |\langle x, \lambda x \rangle|^2 - \langle x, x \rangle \langle \lambda x, \lambda x \rangle \\ 0 &\geq |\lambda|^2 |\langle x, x \rangle|^2 - \left| \lambda^2 \right| (\langle x, x \rangle)^2, \end{aligned}$$

and thus we have equality since $\langle x, x \rangle$ is always real. For the other direction, suppose we have equality. Then in the proof, each inequality becomes an equality, including $0 = \langle x - \alpha y, x - \alpha y \rangle$, which says that $x = \alpha y$, and thus they are scalars of each other. □

Remark 125. The Cauchy-Bunyakowsky-Schwarz Inequality, might be familiar from analysis, where if $x, y \in \mathbb{R}^n$, then

$$\left| \sum_{i=1}^{n} \right| x_i y_i \leq \left(\sum_{i=1}^{n} x_i^2)^{1/2} \right) \left(\sum_{i=1}^{n} y_i^2 \right)^{1/2},$$

where $\|x\| = \left(\sum_{i=1}^{n} x_i^2)^{1/2} \right)$ is the usual Euclidean norm that induces the Euclidean distance.

Next, we show that the collection of inner product spaces is a subspace of the collection of normed linear spaces.

Theorem 126 ([Con90] Theorem 1.1.5)**.** *If $\langle \cdot, \cdot \rangle$ is a inner product on X, and $\|x\| = \langle x, x \rangle^{1/2}$ for all $x \in X$, then $\|\cdot\|$ is a norm on X.*

Proof. We prove the three properties of norms.

(a) Clearly $\|x\| \geq 0$ since $\langle x, y \rangle \geq 0$. Suppose $\|x\| = 0$, then $\langle x, x \rangle^{1/2}$, so $\langle x, x \rangle = 0$, and thus $x = 0$ by the definition of an inner product.

(b) Let $\alpha \in \mathbb{K}$, then
$$\|\alpha x\| = \langle \alpha x, \alpha x \rangle^{1/2} = (|\alpha|^2 \langle x, x \rangle)^{1/2} = |\alpha| \langle x, x \rangle^{1/2} = |\alpha| \|x\|.$$

(c) Let $x, y \in X$, then

$$\begin{aligned} \|x + y\|^2 &= \langle x + y, x + y \rangle \\ &= \|x\|^2 + \langle y, x \rangle + \langle x, y \rangle + \|y\|^2 \\ &= \|x\|^2 + 2 \operatorname{Re} \langle x, y \rangle + \|y\|^2 \end{aligned}$$

where $\langle x, y\rangle + \langle y, x\rangle = \operatorname{Re}\langle x, y\rangle$, since $\langle y, x\rangle = \overline{\langle x, y\rangle}$ and adding a complex number to its conjugate cancels the complex part and leaves twice the real part. By the Cauchy-Bunyakowsky-Schwarz Inequality $\operatorname{Re}\langle x, y\rangle \leq |\langle x, y\rangle| \leq \|x\| \|y\|$. Hence

$$\|x+y\|^2 \leq \|x\| + 2\|x\| \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

taking square roots gives the inequality.

□

Remark 127. We call the identity

$$\|x+y\|^2 = \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y\|^2$$

the *polar identity*.

Remark 128. Now we have the following diagram of implications

$$\text{inner product space} \implies \text{normed space} \implies \text{metric space.}$$

In a metric space we can measure the distance between two elements, and in a normed space we can measure the length of an object. Now, in a inner product space there is a notion of an angle. The inner product space allows us to determine when two elements are orthogonal to each other. We will explore this notion of orthogonality next.

Definition 129. If H is a Hilbert space, and $f, g \in H$, then f and g are *orthogonal* if $\langle f, g\rangle = 0$. We write $f \perp g$. If $A, B \subseteq H$, then $A \perp B$ if $f \perp g$ for all $f \in A$ and $g \in G$.

The following is a generalization of the classical Pythagorean theorem.

Theorem 130 (Pythagorean Theorem, [Con90] Theorem 1.2.2)**.** *If $f_1, \dots, f_n$ are pairwise orthogonal vectors in H, then*

$$\|f_1 + f_2 + \cdots + f_n\|^2 = \|f_1\|^2 + \|f_2\|^2 + \cdots + \|f_n\|^2 .$$

Proof. If $f_1 \perp f_2$, then

$$\|f_1 + f_2\| = \langle f_1 + f_2, f_1 + f_2\rangle = \|f_1\|^2 + 2\operatorname{Re}\langle f_1, f_2\rangle + \|f_2\|^2$$

by the polar identity. Since $f_1 \perp f_2$, $\langle f_1, f_2\rangle = 0$, giving us the identity. The rest of the proof proceeds by induction. □

Theorem 131 (Parallelogram Law, [Con90] Theorem 1.2.3)**.** *If H is a Hilbert space, and $f, g \in H$, then*

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|).$$

Proof. For any $f, g \in H$, the polar identity implies

$$\|f+g\|^2 = \|f\|^2 + 2\operatorname{Re}\langle f, g\rangle + \|g\|^2$$
$$\|f-g\|^2 = \|f\|^2 - 2\operatorname{Re}\langle f, g\rangle + \|g\|^2 .$$

Adding these gives the result. □

There is actually a stronger statement that can be made using the parallelogram law.

Theorem 132. *A norm arises from an inner product if it satisfies the parallelogram law.*

Proof. First assume we are dealing with a vector space X over $\mathbb{R}$. We begin by setting

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

Now, one can check that $\langle x, y\rangle = \langle y, x\rangle$, and that $\|x\| = \sqrt{\langle x, x\rangle}$. The next step is to check that all of the properties of an inner product are satisfied. The first of which is $\langle x+y, x\rangle = \langle x, z\rangle + \langle y, z\rangle$. By the parallelogram law we have

$$2\|x+z\|^2 + 2\|y\|^2 = \|x+y+z\|^2 + \|x-y+z\|^2.$$

This gives

$$\begin{aligned}\|x+y+z\|^2 &= 2\|x+z\|^2 + 2\|y\|^2 - \|x-y+z\|^2 \\ &= 2\|y+z\|^2 + 2\|x\|^2 - \|y-x+z\|^2\end{aligned}$$

where the second formula follows from the first by swapping x and y. This gives

$$\|x+y+z\|^2 = \|x\|^2 + \|y\|^2 + \|x+z\|^2 + \|y+z\|^2 - \frac{1}{2}\|x-y+z\|^2 - \frac{1}{2}\|y-x+z\|.$$

Replacing z by $-z$ in the above equation gives

$$\|x+y-z\|^2 = \|x\|^2 + \|y\|^2 + \|x-z\|^2 + \|y-z\|^2 - \frac{1}{2}\|x-y-z\|^2 - \frac{1}{2}\|y-x-z\|^2.$$

Since $\|-a\| = |-1|\,\|a\| = \|a\|$, we get

$$\begin{aligned}\langle x+y, z\rangle &= \frac{1}{4}\left(\|x+y+z\|^2 - -\|x+y-z\|^2\right) \\ &= \frac{1}{4}(\|x\|^2 + \left\|y^2\right\| + \|x+z\|^2 + \|y+z\|^2 - \frac{1}{2}\|x-y+z\|^2 - \frac{1}{2}\|y-x+z\|^2 \\ &- \|x\|^2 - \|y\|^2 - \|x-z\|^2 - \|y-z\|^2 + \frac{1}{2}\|x-y-z\|^2 + \frac{1}{2}\|y-x-z\|^2) \\ &= \|x+z\|^2 + \|y+z\|^2 - \|x-z\|^2 - \|y-z\|^2 \\ &= \langle x, z\rangle + \langle y, z\rangle.\end{aligned}$$

Next we show that $\langle \lambda x, y\rangle = \langle x, y\rangle$ for all $\lambda \in \mathbb{R}$. First this holds for all $\lambda \in \mathbb{N}$ using the fact that $\langle x+y, z\rangle = \langle x, y\rangle + \langle x, z\rangle$ and induction. Then we check for $\lambda = -1$. Indeed,

$$\begin{aligned}\langle -x, y\rangle &= \frac{1}{4}\left(\|y-x\|^2 - \|-x-y\|^2\right) \\ &= \frac{1}{4}\left(\|x-y\|^2 - \|x+y\|^2\right) \\ &= -\frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right) \\ &= -\langle x, y\rangle.\end{aligned}$$

Thus the formula holds for all $\lambda \in \mathbb{Z}$. If $\lambda = \frac{a}{b} \in \mathbb{Q}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$, we let $x' = \frac{x}{q}$ and thus

$$q\langle \lambda x, y\rangle = q\langle px', y\rangle = p\langle qx', y\rangle = p\langle x, y\rangle,$$

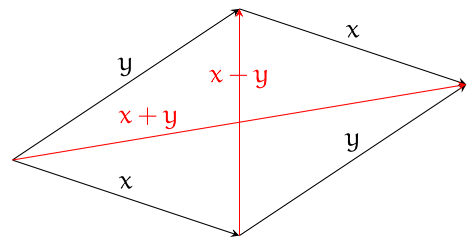


Figure 13.1: Illustration of the parallelogram law

and dividing by q gives

$$\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$$

for all $\lambda \in \mathbb{Q}$. We have proved that for any $x, y \in X$, the continuous function $\lambda \mapsto \frac{1}{\lambda} \langle \lambda x, y \rangle$ defined on $\mathbb{R}\backslash\{0\}$ is equal to $\langle x, y \rangle$ for all $\lambda \in \mathbb{Q}\backslash\{0\}$, thus equality holds for all $t \in \mathbb{R}\backslash\{0\}$, and thus $\langle x, y \rangle$ is indeed an inner product on X.

Now, if X is a vector space over $\mathbb{C}$, we can define

$$\langle x, y \rangle = \frac{1}{4} \sum_{k=0}^{3} i^k \left\| x + i^k y \right\|^2$$

and using the case of real scalars, it can be proven that this is an inner product on X as well. □

There is another similar statement that arises from Euclidean geometry.

Theorem 133 (Ptolemy's Inequality). *Suppose that $\|\cdot\|$ is a norm on a vector space X. Then this norm satisfies Ptolemy's inequality*

$$\|x - y\| \, \|z\| + \|y - z\| \, \|x\| \geq \|x - z\| \, \|y\|$$

for all $x, y, z \in X$ if and only if there exists an inner product $\langle \cdot, \cdot \rangle$ on X such that $\|x\|^2 = \langle x, x \rangle$ for all $x \in X$.

Example. The space $C([a, b], \mathbb{R})$ is a normed space where the norm defined as

$$\|f\| = \max\{f(x) \mid x \in [a, b]\}$$

does not arise from an inner product since this norm does not satisfy the paralellogram inequality. Indeed, take $f(x) = 1$ and $g(x) = \frac{x-a}{b-a}$. We have $\|f\| = 1$ and $\|g\| = 1$ and

$$f(x) + g(x) = 1 + \frac{x - a}{b - a}$$
$$f(x) - g(x) = 1 - \frac{x - a}{b - a}.$$

Hence $\|f + g\| = 2$, $\|f - g\| = 1$ and so $\|f + g\|^2 + \|f - g\|^2 = 5$, but $2(\|f\|^2 + \|g\|^2) = 4$.

Now we turn our attention to linear functionals on Hilbert spaces. Our goal is to prove the Riesz representation theorem which will be crucial when looking at BMO spaces. This section follows a mix of [Con90] and [Fol99].

Definition 134. Let H be a Hilbert space over a field F. A *bounded linear functional* L on H is a linear function $L\colon H \to F$ for which there is a constant $c > 0$ such that $|L(h)| \leq c\,\|h\|$ for all $hinH$. For a bounded linear functional $L\colon H \to F$ define the norm of L to be

$$\|L\| = \sup\{|L(h)| \mid \|h\| \leq 1\}.$$

Example. Let $H = \mathbb{R}^n$, then a linear functional is any linear transformation on $\mathbb{R}^n$. For example, $L(h) = 5h$ is a bounded linear transformation.

The remarkable thing about linear functionals is that continuity and boundedness are actually equivalent conditions.

Theorem 135 ([Con90] Theorem 1.3.1)**.** *If X and Y are normed linear spaces and $L\colon X \to Y$ is a linear functional, the following are equivalent:*

(a) T *is continuous.*

(b) T *is continuous at 0.*

(c) T *is bounded.*

Proof. The fact that (a) implies (b) is immediate. If L is continuous at $0 \in X$, there is a neighborhood U of 0 such that $L(U) \subset \{y \in Y \mid \|y\| \leq 1\}$ and U must contain a ball $B = \{x \in X \mid \|x\| \leq \delta\}$. Thus $\|L(x)\| \leq 1$ when $\|x\| \leq \delta$. Since L commutes with scalar multiplication, it follows that $\|L(x)\| \leq a\delta^{-1}$ whenever $\|x\| \leq a$, that is, $\|L(x)\| \leq \delta^{-1}\,\|x\|$. This show that (b) implies (c).

Finally, if $\|L(x)\| \leq C\,\|x\|$ for all x, then

$$\|L(x_1) - L(x_2)\| = \|L(x_1 - x_2)\| \leq \epsilon$$

whenever $\|x_1 - x_2\| \leq C^1\epsilon$, so that L is continuous. □

Before we can prove the Riesz Representation Theorem we need a couple of facts about orthogonal complements of subspaces. Let H be a Hilbert space and let $A \subset H$. Define

$$A^\perp = \{f \in H \mid f \perp g \text{ for all } g \text{ in } A\}$$

This space is a closed linear subspace of H.

Theorem 136 ([Fol99] Theorem 5.24)**.** *If M is a closed subspace of H, then $H = M \oplus M^\perp$ that is, each $x \in H$ can be expressed uniquely as $x = y + z$ where $y \in M$ and $z \in M^\perp$.*

Proof. Given $x \in H$ let $\delta = \inf\{\|x - y\| \mid y \in M\}$, and let $\{y_n\}$ be a sequence in M such that $\|x - y_n\| \to \delta$. By the parallelogram law,

$$2\left(\|y_n - x\|^2 + \|y_m - x\|^2\right) = \|y_n - y_m\|^2 + \|y_n + y_m - 2x\|^2,$$

so since $\frac{1}{2}(y_n + y_m) \in M$,

$$\begin{aligned}\|y_n - y_m\|^2 &= 2\,\|y_n - x\|^2 + 2\,\|y_m - x\|^2 - 4\left\|\frac{1}{2}(y_n + y_m) - x\right\|^2 \\ &\leq 2\,\|y_n - x\|^2 + 2\,\|y_m - x\|^2 - 4\delta^2.\end{aligned}$$

As $m, n \to \infty$ this last quantity tends to zero, so $\{y_n\}$ is a Cauchy sequence. Let $y = \lim_{n\to\infty} y_n$ and $z = x - y$. Then $y \in M$ since M is closed, and $\|x - y\| = \delta$.

We will now show that $z \in M^{\perp}$. Indeed, if $u \in M$, after multiplying u by a nonzero scalar we may assume that $\langle z, u\rangle$ is real. Then the function

$$f(t) = \|z + tu\|^2 = \|z\|^2 + 2t \langle z, u\rangle + t^2 \|u\|^2$$

is real for $t \in \mathbb{R}$ and it has a minimum at $t = 0$ because $z + tu = x - (y - tu)$ and $(y - tu) \in M$. Thus $2 \langle z, u\rangle = f'(0) = 0$, so $z \in M^{\perp}$. ☐

We begin with a couple of observations. Let H be a Hilbert space over some field F, and fix $h_0 \in H$. Define $L\colon H \to F$ by $L(h) = \langle h, h_0\rangle$. Then L is linear since inner products are linear and the Cauchy-Bunyakowsky-Schwarz Inequality gives

$$|L(h)| = |\langle h, h_0\rangle| \le \|h\| \|h_0\| .$$

This tells us that L is bounded and $\|L\| \le \|h_0\|$. In fact, $L(h_0/\|h_0\|) = \langle h_0/\|h_0\|, h\rangle = \|h_0\|$, so $\|L\| = \|h_0\|$. The Riesz Representation Theorem will give a converse to this.

Theorem 137 ([Fol99] Theorem 5.25)**.** *If* $L\colon H \to F$ *is a bounded linear functional, then there is a unique vector* $h_0 \in H$ *such that* $L(h) = \langle h, h_0\rangle$ *for every* $h \in H$*. Moreover,* $\|L\| = \|h_0\|$.

Proof. First we prove uniqueness. If $\langle h, h_0\rangle = \langle h, h_0'\rangle$ for all $h \in H$, by taking $x = h_0 - h_0'$ we conclude that $\|h_0 - h_0'\|^2$ and hence $h_0 = h_0'$.

Now for the existence. If L is the zero functional, then just take $h_0 = 0$. Otherwise, let $M = \{h \in H \mid L(h) = 0\}$. Then M is a proper closed subspace of H so $M^{\perp} \neq \{0\}$ by Theorem 136. Pick $z \in M^{\perp}$ with $\|z\| = 1$. If $u = L(h)z - L(z)h$ then $u \in M$, so

$$0 = \langle u, z\rangle = L(h) \|z\|^2 - L(z) \langle h, z\rangle = L(h) - \left\langle h, \overline{L(z)} z\right\rangle .$$

Hence $L(h) = \langle h, h_0\rangle$ where $h_0 = \overline{L(z)} z$. ☐

Chapter 14

L^p Spaces

We have examined the spaces $\mathcal{L}^p$ in several examples. Recall that we defined this space to be

$$\mathcal{L}^p(X) = \left\{ f\colon X \to \mathbb{R} \mid \int_X |f(x)|^p \, d\mu < \infty \right\}.$$

We also defined the "norm" on $\mathcal{L}_p$:

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

Using Minkowski's inequality, we know that $\|f+g\| \leq \|f\| + \|g\|$, and we can easily see that $\|\alpha f\| = |\alpha| \, \|f\|$. However, if $\|f\| = 0$, we can only conclude that $f = 0$ almost everywhere, i.e., f could be nonzero on a set of measure zero. An example of this is Dirichlet's function, where $f(x) = 1$ if $x \in \mathbb{Q} \cap [0,1]$ and $f(x) = 0$ if x is irrational. So we introduce an equivalence relation on $\mathcal{L}_p$.

Definition 138. We define $f \sim g$ if $f = g$ almost everywhere.

Proposition 139. *The relation defined above is an equivalence relation.*

Proof. First, we have that $f \sim f$ since $f = f$ everywhere. Next, if $f = g$ almost everywhere, then $g = f$ almost everywhere. Finally, suppose that $f = g$ almost everywhere, and $g = h$ almost everywhere. Let

$$A = \{x \in X \mid f(x) \neq g(x)\}, \quad B = \{x \in X \mid g(x) \neq h(x)\}, \quad C = \{x \in X \mid f(x) \neq h(x)\}.$$

By assumption A and B are sets of measure zero. Then we have that $C \subseteq A \cup B$, but $\mu(A \cup B) \leq \mu(A) + \mu(B) = 0 + 0 = 0$. Thus $\sim$ is an equivalence relation. □

Definition 140. For $1 \leq p < \infty$, we define

$$L^p(X) = \{[f] \mid f \in \mathcal{L}^p\} = \mathcal{L}^p(X)/\sim.$$

We endow L^p with the norm

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

Remark 141. Now, $\|\cdot\|_p$ is a norm, since if $\|f\|_p = 0$, then $f = 0$ almost everywhere, so $[f] = [0]$.

Definition 142. For $p = \infty$, we define L^∞ to be the space of all bounded measurable functions. We define the *essential supremum* to be

$$\|f\|_\infty = \operatorname{ess\,sup}\{|f(x)| \mid x \in X\}$$

where

$$\operatorname{ess\,sup} f(x) = \inf\{M \mid \mu\{x \mid f(x) > M\} = 0\}.$$

Remark 143. The essential supremum takes the smallest number such that the values of f that are greater than it form a set of measure 0. This means that it will be equivalent to a function that is bounded by the essential supremum, since we again identify functions as we did above.

Now we prove that the L^p spaces are Banach spaces for all $1 \le p \le \infty$.

Theorem 144 ([Bas13] Theorem 15.4)**.** *The L^p spaces are complete.*

Proof. We first prove the theorem for $p < \infty$. Suppose f_n is a Cauchy sequence in L^p. We first find a desirable subsequence. Given $\epsilon = 2^{-(j+1)}$, there exists n_j such that if $n, m \ge n_j$, then $\|f_n - f_m\|_p \le 2^{-(j+1)}$. Without loss of generality we may assume that $n_j \ge n_{j-1}$ for each j.

Set $n_0 = 0$ and define $f_0 = 0$. Our candidate for the limit function is $\sum_m (f_{n_m} - f_{n_{m-1}})$. We will now show that this series is absolutely convergent. Set $g_j(x) = \sum_{m=1}^{j} \left|f_{n_m}(x) - f_{n_{m-1}}(x)\right|$. Note that g_j increases in j for each x. Let $g(x)$, which might be infinite, be the limit of the g_j. By Minkowski's inequality

$$\|g_j\|_p \le \sum_{m=1}^{j} \left\|f_{n_m} - f_{n_{m-1}}\right\|_p \le \|f_{n_1} - f_{n_0}\|_p + \sum_{m=2}^{j} 2^{-m} \le \|f_{n_1}\|_p + \frac{1}{2}.$$

By Fatou's lemma,

$$\int_X |g(x)|^p \, d\mu \le \lim_{j\to\infty} \int \left|g_j(x)\right|^p d\mu = \lim_{j\to\infty} \|g_j\|_p^p \, d\mu \le \frac{1}{2} \|f_{n_1}\|_p .$$

Hence g is finite almost everywhere. this prove the absolute convergence for almost every x.

Now we define our function f. Set

$$f(x) = \sum_{m=1}^{\infty} \left(f_{n_m}(x) - f_{n_{m-1}}(x)\right).$$

We just showed that this series is absolutely convergent for almost every x. Set $f(x) = 0$ for any x where absolute convergence does not hold. Note that since the set of x where the function is not absolutely convergent is a set of measure 0, defining $f(x) = 0$ on these points does not change the equivalence class of the function in L^p. Now we have

$$f(x) = \lim_{K\to\infty} \sum_{m=1}^{K} \left(f_{n_m}(x) - f_{n_{m-1}}(x)\right) = \lim_{K\to\infty} f_{n_K}(x)$$

since the series is telescoping. By Fatous's lemma

$$\left\|f - f_{n_j}\right\|_p^p = \int |f - f_{n_k}|^p \, d\mu \le \liminf_{K\to\infty} \int \left|f_{n_K} - f_{n_j}\right|^p = \liminf_{k\to\infty} \left\|f_{n_K} - f_{n_j}\right\|_p^p \le 2^{-(j+1)p}.$$

We have thus shown that $\left\|f - f_{n_j}\right\|_p \to 0$ as $j \to \infty$. Now, given $\epsilon > 0$, there exists N such that $\|f_n - f_m\|_p \le \epsilon$ if $n, m > N$. In particular, $\left\|f_{n_j}\right\|_p < \epsilon$ if j is large enough. By Fatou's lemma

$$\|f - f_m\|_p^p \le \liminf_{j\to\infty} \left\|f_{n_j} - f_m\right\|_p^p \le \epsilon^p$$

if $m \ge N$. This shows that f_m converges to f in the L^p norm.

Finally we show the case when $p = \infty$. Suppose f_n is a Cauchy sequence in L^∞. Let $A_k = \{x \mid |f_k(x)| \ge \|f_k\|_\infty\}$ and let $B_{m,n} = \{x \mid |f_n(x) - f_m(x)| \ge \|f_n - f_m\|_\infty\}$. Let E be the union of these sets for $k, m, n = 1, 2, \ldots$. Then $\mu(E) = 0$, and on the complement of E the sequence f_n converges uniformly to a bounded function f. Define $f(x) = 0$ for $x \in E$. Then $f \in L^\infty$ and $\|f - f_n\|_\infty \to 0$ as $n \to \infty$. □

Now we can actually show that the L^p spaces are not just any complete space, but they are actually the completion of the spaces $C([a, b], \mathbb{R})$. We will prove something slightly more general. We say that a function f is *compactly supported*, if $\mu(\{x \in X \mid f(x) \ne 0\}) < \infty$, i.e., the subset of the domain of f which is not in the kernel of f, has finite measure. We denote the set of continuous, compactly supported functions on X by $C_c(X)$.

Now we briefly discuss linear functionals on L^p spaces. Our ultimate goal is to prove that the dual space of L^p is L^q where $p^{-1} + q^{-1} = 1$. This section follows the book [Bas13]

Definition 145. A *linear functional* on a normed linear space $(X, \|\cdot\|)$ is a mapping, F, of X into $\mathbb{R}$ such that $F(\alpha f + \beta g) = \alpha F(f) + \beta F(g)$. A linear functional is *bounded* if there is a constant M such that $|F(f)| \le M \cdot \|f\|$ for all $f \in X$. The smallest constant for which this is true is called the *norm* of F. Thus we have

$$\|F\| = \sup\left\{\frac{|F(f)|}{\|f\|} \mid f \in X\right\}.$$

Example. Define $F\colon L^p \to \mathbb{R}$ by $F(f) = f(0)$. We claim that F is a linear functional. Indeed,

$$F(f + g) = (f + g)(0) = f(0) + g(0) = F(f) + F(g)$$

and if $\alpha \in \mathbb{R}$, then

$$F(\alpha f) = \alpha f(0) = \alpha F(f).$$

Thus F is a linear functional.

Before we will proceed we quickly define the sign function by

$$\operatorname{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0. \end{cases}$$

Note that $|x| = x \operatorname{sgn}(x)$. This function will prove very useful in the upcoming theorems.

Theorem 146 ([Bas13] Theorem 15.9)**.** *For* $1 < p < \infty$ *and* $p^{-1} + q^{-1} = 1$,

$$\|f\|_p = \sup\left\{\int fg\,d\mu \mid \|g\|_q \le 1\right\}.$$

When $p = 1$ *this also holds if* $q = \infty$*, similarly, if* $p = \infty$ *we take* $q = 1$.

Proof. The right hand side of the equation is less than or equal to the left hand side by Hölder's inequality. So we show the other direction. The proof is broken up into 3 cases.

First, suppose $p = 1$. Take $g(x) = \text{sgn}(f(x))$. Then $|g|$ is bounded by 1 and $fg = |f|$. Thus the statement is proven for $p = 1$.

Next, suppose $p = \infty$. If $\|f\|_\infty = 0$ then the statement is trivial, so suppose $\|f\|_\infty > 0$. There exists a sequence of sets F_n such that each $\mu(F_n) < \infty$ and $\mathbb{R}^n = \bigcup_{n=1}^\infty F_n$. If $M = \|f\|_\infty$, let $a < M$ be any finite real number. By the definition of the L^∞ norm, the measure of $A_n = \{x \in F_n \mid |f(x)| > a\}$ must be positive if n is sufficiently large, otherwise the $\|f\|_\infty \leq a$. Let

$$g_n(x) = \frac{\text{sgn}(f(x))\chi_{A_n}(x)}{\mu(A_n)}.$$

Then the L^1 norm of g_n is 1 and

$$\int fg d\mu = \int_{A_n} \frac{|f|}{\mu(A_n)} dx \geq a.$$

Since a is arbitrary, the supremum of the right hand side must be M.

Finally suppose $1 < p < \infty$. We may suppose $\|f\|_p > 0$. Let F_n be the measurable sets of finite measure increasing to $\mathbb{R}^n$, q_n a sequence of non-negative simple functions increasing to f^+, r_n a sequence of non-negative simple functions increasing to f^-, and

$$s_n(x) = (q_n(x) - r_n(x))\chi_{F_n(x)}.$$

Then $s_n(x) \to f(x)$ for all x, $|s_n(x)|$ increases to $|f(x)|$ for each x, each s_n is a simple function, and $\|s_n\|_p < \infty$ for each n. Then $\|s_n\|_p \to \|f\|_p$ by the monotone convergence theorem, whether or not $\|f_p\|$ is finite. For n sufficiently large, $\|s_n\|_p > 0$. Let

$$g_n(x) = (\text{sgn}(f(x)))\frac{|s_n(x)|^{p-1}}{\|s_n\|_p^{p/q}}.$$

Then g_n is a simple function. Since $(p-1)/q = p$ we have

$$\|g_n\|_q = \frac{\left(\int |s_n|^{(p-1)q}\right)^{1/q}}{\|s_n\|_p^{p/q}} = \frac{\|s_n\|_p^{p/q}}{\|s_n\|_p^{p/q}} = 1.$$

On the other hand, since $|f| \geq |s_n|$,

$$\int fg_n = \frac{\int |f|\,|s_n|^{p-1}}{\|s_n\|_p^{p/q}} \geq \frac{\int |s_n|^p}{\|s_n\|_p^{p/q}} = \|s_n\|_p^{p-(p/q)}.$$

Since $p = p/q = 1$ we obtain $\int fg_n \geq \|s_n\|_p$, which tends to $\|f\|_p$. This prove the right hand side of the equation is at least as large as the left hand side. □

From this we also immediately obtain the following corollary.

Corollary 147 ([Bas13] Corollary 15.10)**.** *For* $1 < p < \infty$ *and* $p^{-1} + q^{-1} = 1$,

$$\|f\|_p = \sup\left\{\int fg \mid \|g\|_q \leq 1, g \textit{ simple}\right\}.$$

To finish the proof that $(L^p)^* = L^q$ we need to prove that a function defined by $H(f) = \int fg$ for some $g \in L^q$ and all $f \in L^p$ is a bounded linear functional, and then conversely we need to show that each bounded linear functional on L^p can be represented in such a form.

Proposition 148 ([Bas13] Theorem 15.11)**.** *Suppose* $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, *ad* $g \in L^q$. *If we define* $H(f) = \int fg$ *for* $f \in L^p$, *then* H *is a bounded linear functional on* L^p *ad* $\|H\| = \|g\|_q$.

Proof. Linearity follows from basic properties of Lebesgue integrals. The fact that $\|H\| \leq \|g\|_q$ follows directly from Hölder's inequality. Using Theorem 146 and writing

$$\|H\| \sup_{\|f\|_p \leq 1} |H(f)| = \sup_{\|f\|_p} \left| \int fg \right| \geq \sup_{\|f\|_p} \int fg = \|g\|_q$$

completes the proof. □

The last piece is the following.

Theorem 149 ([Bas13] Theorem 15.12)**.** *Suppose* $1, p < \infty, p^{-1} + q^{-1} = 1$, *and* H *is a real-valued bounded linear functional on* L^p. *Then there exists* $g \in L^q$ *such that* $H(f) = \int fg$ *an* $\|g\|_q = \|H\|$.

Proof. Suppose we are given a bounded linear functional H on $L^p(X)$. First assume that $X \subset \mathbb{R}^n$ such that $\mu(X) < \infty$. Define $\nu(A) = H(\chi_A)$. We will show that ν is a measure, that $\nu \ll \mu$ and that $g = d\nu/d\mu$ is the function we seek.

If A and B are disjoint, then

$$\nu(A \cup B) = H(\chi_{A \cup B}) = H(\chi_A + \chi_B) = H(\chi_A) + H(\chi_B) = \nu(A) + \nu(B).$$

To show that ν is countably additive, it suffices to show that if $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \bigcup_{n=1}^\infty A_n$, then $\nu(A_n) \to \nu(A)$. But if the A_n are such a sequence of sets, then $\chi_{A_n} \to \chi_A$ in L^p, and so $\nu(A_n) = H(\chi_{A_n}) \to H(\chi_A) = \nu(A)$. We use here the fact that $\mu(X) < \infty$. We can conclude that ν is a countable additive signed measure. Moreover, if $\mu(A) = 0$, then $\chi_A = 0$ almost everywhere, hence $\nu(A) = H(\chi_A) = 0$. Using the Radon-Nikodym theorem for signed measures, we see there exists a real-valued integrable function g such that $\nu(A) = \int_A g$ for all sets A.

If $s = \sum_i a_i \chi_{A_i}$ is a simple function, by linearity we have

$$H(s) = \sum_i a_i H(\chi_{A_i}) = \sum_i a_i \nu(A_i) = \sum_i a_i \int g \chi_{a_i} = \int gs.$$

By Corollary 147

$$\|g\|_q = \sup \left\{ \int gs \mid \|s\|_p \leq 1, s \text{ simple} \right\} = \sup \left\{ H(s) \mid \|s\|_p \leq 1, s \text{ simple} \right\} \leq \|H\|.$$

If s_n are simple functions tending to f in L^p then $H(s_n) \to H(f)$, while by Hölder's inequality

$$\left| \int s_n g - \int fg \right| = \left| \int (s_n - f) g \right| \leq \|s_n - f\|_p \|g\|_q \to 0,$$

so $\int s_n g \to \int fg$. We thus have $H(f) = \int fg$ for all $f \in L^p$, and $\|g\|_q \leq \|H\|$. By Hölder's inequality $\|H\| \leq \|g\|_q$.

Now suppose that X is σ-finite. Let $F_1 \subseteq F_2 \subseteq \cdots$ so that $\bigcup_{i=1}^\infty F_i = X$ and such that $\mu(F_i) < \infty$. Define functionals H_n by $H_n(f) = H(f\chi_{F_n})$. Clearly each H_n is a bounded linear functional on L^p. Applying the argument above, we see that there exists a g_n such that $H_n(f) = \int f g_n$ and $\|g_n\|_q = \|H_n\| \leq \|H\|$. It is easy to see that g_n is 0 if $x \notin F_n$. Moreover, by the uniqueness part of the Radon-Nikodym theorem, if $n > m$, then $g_n = g_m$ on F_m. Define g by setting $g(x) = g_n(x)$ if $x \in F_n$. Then g is well-defined. By Fatou's lemma, g is in L^q with a norm bounded by $\|H\|$. Note that $f\chi_{F_n} \to f$ in L^p by the dominated convergence theorem. Since H is a bounded linear functional on L^p, we have $H_n(f) = H(f\chi_{F_n}) \to H(f)$. On the other hand

$$H_n(f) = \int_{F_n} f g_n = \int_{F_n} f g \to \int f g$$

by the dominated convergence theorem. Thus $H(f) = \int f g$. Again by Hölder's inequality $\|H\| \leq \|g\|_q$. □

Chapter 15

Hardy-Littlewood Maximal Operator

We begin with a preliminary section, before we embark on the big goal of BMO spaces. This section will provide both background and motivation. However, before we get to this we make one more detour to distribution functions and weak L^p spaces.

15.1 Distribution Functions and Weak L^p

Definition 150 (Distribution Function)**.** For f a measurable function on X, the *distribution function* of f is the function d_f defined on $[0, \infty)$ as follows:

$$d_f(\alpha) = \mu(\{x \in X \mid |f(x)| > \alpha\}).$$

The distribution function captures the size of f, but not the behavior of f around any particular point. Note that the distribution function is nonincreasing. The decrease of $d_f(\alpha)$ as α grows describes the relative largeness of the function, while the increase of $d_f(\alpha)$ as α tends to zero describes the relative smallness of the function at infinity.

Example.

(a) Let

$$f(x) = \sum_{j=1}^{N} a_j \chi_{E_j}(x)$$

be a simple function where the E_j are pairwise disjoint and $a_1 > \cdots > a_N$. If $\alpha \geq a_1$, then $d_f(\alpha) = 0$. However, if $a_2 \leq \alpha < a_1$, then $|f(x)| > \alpha$ when $x \in E_1$, and in general, if $a_{j+1} \leq \alpha < a_j$, then $|f(x)| > \alpha$ precisely when $x \in \bigcup_{k=1}^{j} E_k$. Setting

$$B_j = \sum_{k=1}^{j} \mu(E_k),$$

for $j \in \{1, \ldots, N\}$, $B_0 = a_{N+1} = 0$ and $a_0 = \infty$, we have

$$d_f(\alpha) = \sum_{j=0}^{N} B_j \chi_{[a_{j+1}, a_j)}(\alpha).$$

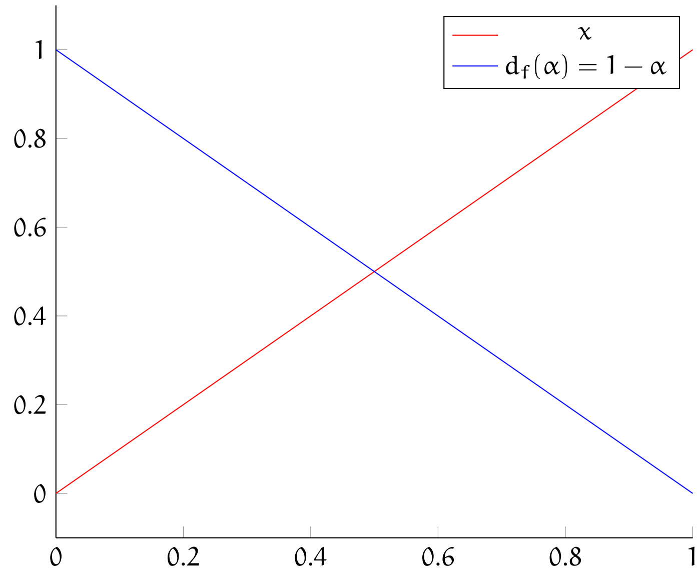


Figure 15.1: Graph of $f(x) = x$ and $d_f(\alpha) = 1 - \alpha$.

(b) Consider the function $f(x) = x\chi_{[0,1]}$. If $\alpha < 0$, then $d_f(\alpha) = \infty$. Suppose that $\alpha \in [0, 1]$. Then $\mu(\{x \in X \mid |x| > \alpha\}) = \mu((\alpha, 1)) = 1 - \alpha$. Finally, if $\alpha > 1$, then $d_f(\alpha) = 0$. See figure 15.1.

(c) Next, consider the function $f(x) = x^2\chi_{[0,1]}$. If $\alpha \in [0, 1]$ then

$$d_f(\alpha) = \mu(\{x \in X \mid x^2 > \alpha\}) = \mu((\sqrt{\alpha}, 1)) = 1 - \sqrt{\alpha}.$$

And if $\alpha > 1$, then $d_f(\alpha) = 0$. See figure 15.2.

Proposition 151 (Properties of d_f, [Gra14a] Proposition 1.1.3)**.** *Let f and g be measurable functions on (X, μ). Then for all $\alpha, \beta > 0$ we have*

(a) $|g| \leq |f|$ *μ-a.e. implies that* $d_g \leq d_f$.

(b) $d_{cf}(\alpha) = d_f(\alpha/|c|)$, *for all* $c \in \mathbb{C}\backslash\{0\}$.

(c) $d_{f+g}(\alpha + \beta) \leq d_f(\alpha) + d_g(\beta)$.

(d) $d_{fg}(\alpha\beta) \leq d_f(\alpha) + d_g(\beta)$.

Proof. For (a) let $\alpha \in X$, then $d_g(\alpha) = \mu(\{x \in X\colon |g(x)| > \alpha\})$ but since $\alpha < |g(x)| < |f(x)|$ almost everywhere,

$$d_g(\alpha) \leq \mu(\{x \in X\colon |f(x)| > \alpha\}) = d_f(\alpha).$$

Next, for (b) we have

$$d_{cf}(\alpha) = \mu(\{x \in X\colon |cf(x)| > \alpha\}) \tag{15.1}$$
$$= \mu(\{x \in X\colon |f(x)| > \alpha/|c|\}) \tag{15.2}$$
$$= d_f(\alpha/|c|). \tag{15.3}$$

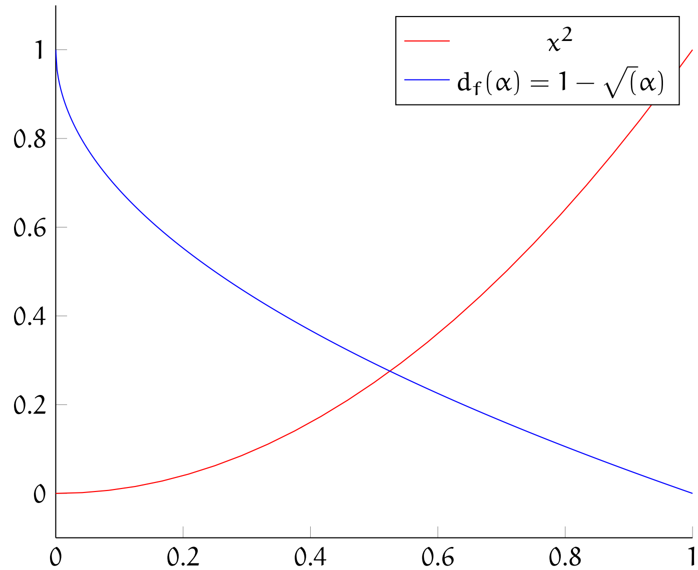


Figure 15.2: Graph of $f(x) = x$ and $d_f(\alpha) = 1 - \alpha$.

To prove (c) we have

$$\begin{aligned} d_{f+g}(\alpha+\beta) =& \mu\left(\{x \in X \colon |f(x)+g(x)| > \alpha+\beta\}\right) & (15.4)\\ \leq& \mu\left(\{x \in X \colon |f(x)|+|g(x)| > \alpha+\beta\}\right) & (15.5)\\ \leq& \mu\left(\{x \in X \colon |f(x)| > \alpha\}\right) + \mu\left(\{x \in X \colon |g(x)| > \beta\}\right) = d_f(\alpha) + d_g(\beta), & (15.6) \end{aligned}$$

where the last inequality comes from the fact that there could exist an $x \in X$ such that either $|f(x)| > \alpha$ or $|g(x)| > \beta$, but $|f(x)| + |g(x)| \leq \alpha + \beta$, but if $|f(x)| + |g(x)| > \alpha + \beta$, then either $|f(x) > \alpha$ or $|g(x)| > \beta$, so the x will still be accounted for in the second term. Finally, for (d) we have

$$\begin{aligned} d_{f+g}(\alpha+\beta) =& \mu\left(\{x \in X \colon |f(x)g(x)| > \alpha+\beta\}\right) & (15.7)\\ =& \mu\left(\{x \in X \colon |f(x)||g(x)| > \alpha+\beta\}\right) & (15.8)\\ \leq& \mu\left(\{x \in X \colon |f(x)| > \alpha\}\right) \cdot \mu\left(\{x \in X \colon |g(x)| > \beta\}\right) = d_f(\alpha)d_g(\beta), & (15.9) \end{aligned}$$

by the same reasoning as in (c). □

Now we will show that we can actually use the distribution function to calculate the L^p norm of f.

Proposition 152 ([Gra14a] Proposition 1.1.4)**.** *Let (X, μ) be a σ-finite measure space. Then for $f \in L^p(X, \mu)$, $0 < p < \infty$, we have*

$$\|f\|_{L^p}^p = p \int_0^\infty \alpha^{p-1} d_f(\alpha)\, d\alpha.$$

Moreover, for any increasing continuously differentiable function φ on $[0, \infty)$ with $\varphi(0) = 0$ and every measurable function f on X with $\varphi(|f|)$ integrable on X, we have

$$\int_X \varphi(|f|)\, d\mu = \int_0^\infty \varphi'(\alpha) d_f(\alpha)\, d\alpha.$$

Proof. First we have

$$p\int_0^\infty \alpha^{p-1} d_f(\alpha)d\alpha = p\int_0^\infty \alpha^{p-1}\int_X \chi_{\{x:\,|f(x)|>\alpha\}} d\mu(x)d\alpha \tag{15.10}$$

$$= \int_x \int_0^{|f(x)|} p\alpha^{p-1} d\alpha d\mu(x) \tag{15.11}$$

$$= \int_X |f(x)|^p d\mu(x) \tag{15.12}$$

$$= \|f\|_{L^p}^p , \tag{15.13}$$

where the second inequality comes from Fubini's theorem, which requires the measure space to be σ-finite. For the second equality we have

$$\int_0^\infty \varphi'(\alpha)d_f(\alpha)d\alpha = \int_0^\infty \varphi'(\alpha)\int_X \chi_{\{x:\,|f(x)|>\alpha\}} d\mu(x)d\alpha \tag{15.14}$$

$$= \int_X \int_0^{|f(x)|} \varphi'(\alpha)d\alpha d\mu(x) \tag{15.15}$$

$$= \int_X \varphi(|f(x)|)d\mu(x). \tag{15.16}$$

□

Definition 153 (Weak L^p)**.** For $0 < p < \infty$, the space *weak* $L^p(X, \mu)$ is defined as the set of all μ-meausurable functions f such that

$$\|f\|_{L^{p,\infty}} = \inf\{C > 0 \mid d_f(\alpha) \le \frac{C^p}{\alpha^p} \text{ for all } \alpha > 0\} \tag{15.17}$$

$$= \sup\{\gamma d_f(\gamma)^{1/p} \mid \gamma > 0\} \tag{15.18}$$

is finite, and where d_f is the distribution function of f. The space *weak* $L^\infty(X, \mu)$ is by definition $L^\infty(X, \mu)$. Note that this is not actually a normed space though. Indeed, if $f = 0$ almost everywhere, then $\|f\|_{L^{p,\infty}} = 0$, so there can be functions that are not zero everywhere that have zero norm.

We quickly show that the two definitions of $\|f\|_{L^{p,\infty}}$ are equal. We have

$$\inf\{C > 0 \mid d_f(\alpha) \le \frac{C^p}{\alpha^p} \text{ for all } \alpha > 0\} = \inf\{C > 0 \mid \alpha d_f(\alpha)^{1/p} \le C \text{ for all } \alpha > 0\}.$$

Setting $C = \sup\{\gamma d_f(\gamma)^{1/p} \mid \gamma > 0\}$, we clearly have that $\alpha d_f(\alpha)^{1/p} \le C$ for all α. So the two are in fact equal.

Remark 154. The norm we defined above is not actually a norm. It is only a semi-norm.

Proposition 155 ([Gra14a] Proposition 1.1.6)**.** *($L^p(X, \mu) \subseteq L^{p,\infty}$) For any $0 < p < \infty$ and any $f \in L^p(X, \mu)$ we have*

$$\|f\|_{L^{p,\infty}} \le \|f\|_{L^p} .$$

Hence the embedding $L^p(X, \mu) \subseteq L^{p,\infty}(X, \mu)$ holds.

Proof. We apply Chebyshev's Inequality with the function $g(x) = |x|$:

$$d_f(\alpha) = \mu(\{x \in X \mid |f(x)| \geq \alpha\}) \leq \frac{1}{\alpha^p} \int_{|f(x)|>\alpha} |f(x)|^p \, d\mu \leq \frac{1}{\alpha^p} \int_X |f(x)|^p \, d\mu = \|f\|_{L^p}^p$$

Then

$$\|f\|_{L^{p,\infty}} = \sup\{\alpha d_f(\alpha)^{1/p} \mid \alpha > 0\} \leq \sup \left(\int_{|f|>\alpha} |f|^p \, d\mu \right)^{1/p} \leq \left(\int_X |f|^p \, d\mu \right)^{1/p} = \|f\|_{L^p}.$$

□

15.2 Hardy-Littlewood Maximal Operator

One might ask why we are always taking the supremum of functions over a certain set, such as cubes or balls in $\mathbb{R}^n$. Why not the limit as the diameter of the sets go to 0? We start with some motivation. First consider the differentiation of the integral for one dimensional functions. Suppose f is given on $[a, b]$ and it is integrable on that interval, we let

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

for $x \in [a, b]$. Now we can look at $F'(x)$ using the quotient definition of the derivative

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}.$$

Notice that this can be rewritten as

$$\frac{1}{h} \int_x^{x+h} f(y) dy = \frac{1}{|I|} \int_I f(y) dy,$$

where $I = (x, x+h)$, and $|I|$ is the length of this interval. Now, one might see that this is the average of f over I. As $|I| \to 0$ we might expect that

$$\lim_{\substack{|I| \to 0 \\ x \in I}} \frac{1}{|I|} \int_I f(y) dy = f(x)$$

holds for suitable x. In higher dimensions we can reformulate this as the question: does

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

for almost all x? The above equation is true if f is continuous at x. Now, in order to study the limit, it is often convenient to replace $\lim_{r\to 0}$ with $\sup_{r>0}$ to see what the maximal value that an integral of this form can take is.

Given a Lebesgue measurable subset $A \subseteq \mathbb{R}^n$, we denote by $|A|$ its Lebesgue measure. Given $\delta > 0$ and a locally integrable function f on $\mathbb{R}^n$, let

$$Avg_{B(x,\delta)}|f| = \frac{1}{|B(x,\delta)|} \int_{B(x,\delta)} |f(y)| dy$$

denote the average of $|f|$ over the ball of radius δ centered at x.

Definition 156 (Centered Hardy-Littlewood maximal function)**.** Let f be a locally integrable function on $\mathbb{R}^n$. The function

$$\mathcal{M}(f)(x) = \sup_{\delta>0} Avg_{B(x,\delta)}|f| = \sup_{\delta>0} \frac{1}{v_n \delta^n} \int_{|y|<\delta} |f(x-y)|dy,$$

where v_n is the Lebesgue measure of the unit ball in $\mathbb{R}^n$, is called the *centered Hardy-Littlewood maximal function* of f. We change the area we integrate over to all y such that $|y| < \delta$, so that we can work with balls centered at the origin, this creates the change of variables $f(y)$ to $f(x-y)$.

Example. On $\mathbb{R}$, let f be the characteristic function of the interval $[a, b]$. For $x \in (a, b)$, clearly $\mathcal{M}(f) = 1$. For $x \geq b$ it can be seen that the largest average of f over all intervals $(x-\delta, x+\delta)$ is obtained when $\delta = x - a$, so that the whole interval where f is nonzero is included. Similarly, if $x \leq a$, the largest average is obtained when $\delta = b - x$. Therefore,

$$\mathcal{M}(f)(x) = \begin{cases} (b-a)/(2|x-b|) & \text{when } x \leq a, \\ 1 & \text{when } x \in (a,b), \\ (b-a)/(2|x-a|) & \end{cases}$$

Note that $\mathcal{M}$ is a sublinear operator, i.e., $\mathcal{M}(f+g) \leq \mathcal{M}(f) + \mathcal{M}(g)$ for all locally integrable functions f and g. Indeed

$$\mathcal{M}(f+g) = \sup_{\delta>0} \frac{1}{\delta^n v_n} \int_{|y|<\delta} |f(x-y) + g(x-y)|dy \tag{15.19}$$

$$\leq \sup_{\delta>0} \frac{1}{\delta^n v_n} \int_{|y|<\delta} |f(x-y)|dx + \frac{1}{\delta^n v_n} \int_{|y|<\delta} |g(x-y)|dy \tag{15.20}$$

$$\leq \sup_{\delta>0} \frac{1}{\delta^n v_n} \int_{|y|<\delta} |f(x-y)|dx + \sup_{\delta>0} \frac{1}{\delta^n v_n} \int_{|y|<\delta} |g(x-y)|dx \tag{15.21}$$

$$= \mathcal{M}(f) + \mathcal{M}(g). \tag{15.22}$$

Also, $\mathcal{M}(\lambda f) = |\lambda| \mathcal{M}(f)$ for all complex constants λ. The following is an interesting fact about $\mathcal{M}$.

Theorem 157. *If $f \in L^1(\mathbb{R}^n)$ is not identically zero, then $\mathcal{M}(f)$ is never integrable on the whole of $\mathbb{R}^n$, i.e., $\mathcal{M}(f) \notin L^1(\mathbb{R}^n)$.*

Proof. We can choose an N large enough such that

$$\int_{B(0,n)} |f(y)dy \geq \frac{1}{2} \|f\|_{L^1}.$$

Then, we take an $x \in \mathbb{R}^n$ such that $|x| \geq N$. Let $r = |x| + N$, we have

$$\mathcal{M}(f) \geq \frac{1}{|B(x,r)} \int_{B(x,r)} |f(y)|dy \tag{15.23}$$

$$= \frac{1}{v_n(|x|+N)^n} \int_{B(x,r)} |f(y)|dy \tag{15.24}$$

$$\geq \frac{1}{v_n(|x|+N)^n} \int_{B(0,N)} |f(y)|dy \tag{15.25}$$

where the last inequality comes from the fact that $B(0,N) \subset B(x,r)$. Then

$$\frac{1}{\nu_n(|x|+N)^n}\int_{B(0,N)}|f(y)|dy \geq \frac{1}{2\nu_n(|x|+N)^n}\|f\|_{L^1} \tag{15.26}$$

$$\geq \frac{1}{2\nu_n(2|x|)^n}\|f\|_{L^1}. \tag{15.27}$$

It follows that for sufficiently large $|x|$, we have

$$\mathcal{M}(f)(x) \geq c|x|^{-n}$$

where $c = (\nu_n 2^{n+1})^{-1}\|f\|_{L^1}$. It follows that if we integrated the right hand side of this over $\mathbb{R}^n$ it would be infinite, and thus $\mathcal{M}(f)(x) \notin L^1(\mathbb{R}^n)$ unless $\|f\|_{L^1} = 0$, which implies that $f = 0$ almost everywhere. □

Corollary 158. *If $\mathcal{M}(f)(x) = 0$ for some $x_0 \in \mathbb{R}^n$, then $f = 0$ almost everywhere.*

Proof. Take $x = x_0$ in the previous theorem, and we get

$$0 = \mathcal{M}(f)(x) \geq \frac{1}{\nu_n(|x|+N)^n}\int_{B(0,N)}|f(y)|dy$$

and thus

$$0 = \int_{B(0,N)}|f(y)|dy$$

for all sufficiently large N. Thus $f = 0$ almost everywhere. □

Definition 159 (Uncentered Hardy-Littlewood Maximal Function)**.** The *uncentered Hardy-Littlewood maximal function* of f,

$$M(f)(x) = \sup_{\substack{\delta>0 \\ |y-x|<\delta}} \mathrm{Avg}_{B(y,\delta)}|f|,$$

is defined to be the supremum of the averages of $|f|$ over all open balls $B(y,\delta)$ that contain the point x.

Note that $\mathcal{M}(f) \leq M(f)$ for the simple fact that we are taking the supremum of a larger set of balls in the uncentered version than in the centered one.

Example. Let f be the characteristic function on the interval $I = [a,b]$. For $x \in (a,b)$ we have $M(f)(x) = 1$, but if $x \geq b$ we can see that the largest average of f over all intervals $(y-\delta, y+\delta)$ occurs when $\delta = \frac{1}{2}(x-a)$ and $y = \frac{1}{2}(x+a)$. Similarly, when $x \leq a$, the largest average obtained is when $\delta = \frac{1}{2}(b-x)$ and $y = \frac{1}{2}(b+x)$. We can conclude that

$$M(f)(x) = \begin{cases} (b-a)/|x-b| & \text{when } x \leq a \\ 1 & \text{when } x \in (a,b) \\ (b-a)/|x-a| & \text{when } x \geq b. \end{cases}$$

Note that $M(f)$ does not have a jump at $x = a$ and $x = b$, while $\mathcal{M}(f)$ did have jumps at those points.

Theorem 160 ([Gra14a] Theorem 2.1.6)**.** *The uncentered and centered Hardy-Littlewood maximal operators M and $\mathcal{M}$ map $L^1(\mathbb{R}^n)$ to $L^{1,\infty}(\mathbb{R}^n)$, where $L^{p,\infty}$ is weak L^p, with constant at most 3^n and also $L^p(\mathbb{R}^n)$ to $L^p(\mathbb{R}^n)$ for $1 < p < \infty$ with constant at most $3^{n/p}p(p-1)^{-1}$. For any $f \in L^1(\mathbb{R}^n)$ we also have*

$$|\{M(f) > \alpha\}| \leq \frac{3^n}{\alpha}\int_{\{M(f)>\alpha\}}|f(y)|dy.$$

The idea of the proof uses a covering lemma for balls which we will not prove here.

Chapter 16

BMO Spaces

In this section we will look at spaces of bounded mean oscillation (BMO). Loosely, a function is in BMO if it doesn't grow too quickly. For example, the function $\log|x|$ is in BMO while e^x is not. They were initially introduced by Fritz John and Louis Nirenberg for the purpose of studying PDEs. About a decade after their introduction Charles Fefferman proved that the dual space of the Hardy space H^1 is BMO, answering a long standing open problem. We will introduce these spaces, and one of the most important theorems about them, the John-Nirenberg Theorem. To do this we will also look at the Calderón-Zygmund decomposition.

This section is based off of [Gra14b].

Definition 161. Let f be a locally integrable function on $\mathbb{R}^n$ and Q a measurable set in $\mathbb{R}^n$. Denote by

$$\operatorname{Avg}_Q f = \frac{1}{|Q|}\int_Q f(x)dx$$

the *mean of* f *over* Q, where $|Q|$ is the Lebesgue measure of Q. Then the *oscillation* of f over Q is the function $\left|f - \operatorname{Avg}_Q f\right|$, and the *mean oscillation* of f over Q is

$$\frac{1}{|Q|}\int_Q \left|f(x) - \operatorname{Avg}_Q f\right| dx.$$

This quantity measures how far a function gets from its average, on average.

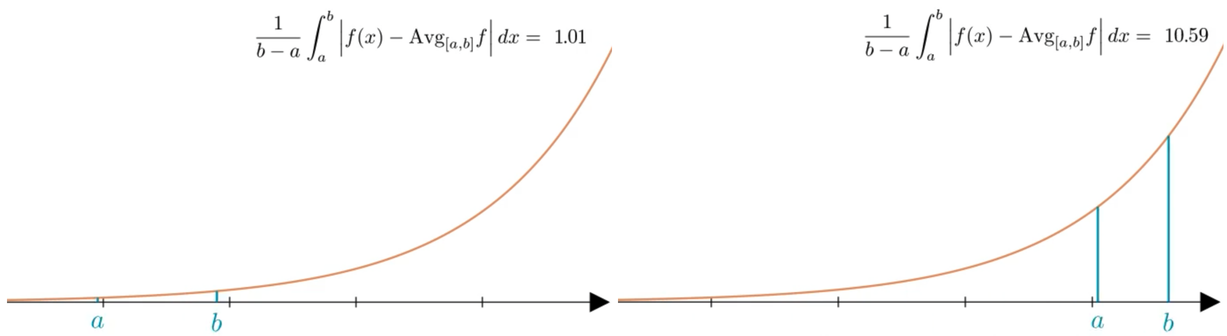
Definition 162. For f a complex-valued locally integrable function on $\mathbb{R}^n$, set

$$\|f\|_{BMO} = \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) - \operatorname{Avg}_Q f\right| dx \mid Q \text{ a cube in } \mathbb{R}^n\right\}.$$

We say that f has *bounded mean oscillation* if $\|f\|_{BMO} < \infty$ and $BMO(\mathbb{R}^n)$ is the set of all locally integrable functions f on $\mathbb{R}^n$ with $\|f\|_{BMO} < \infty$.

Example. Here we see that e^x is not in BMO. The mean oscillation continues to increase as x increases and because of this it cannot be in BMO.

Proposition 163. $BMO(\mathbb{R}^n)$ *is a linear space.*



Proof. Let $\alpha \in \mathbb{R}, f \in \mathrm{BMO}(\mathbb{R}^n)$. Then

$$\begin{aligned} \|\alpha f\|_{BMO} &= \sup\left\{\frac{1}{|Q|}\int_Q \left|\alpha f(x) - \mathrm{Avg}_Q\, \alpha f\right| dx \mid Q \text{ a cube in } \mathbb{R}^n\right\} \\ &= \sup\left\{\frac{1}{|Q|}\int_Q \left|\alpha f(x) - \alpha\, \mathrm{Avg}_Q\, f\right| dx \mid Q \text{ a cube in } \mathbb{R}^n\right\} \\ &= |\alpha|\, \|f\|_{BMO} < \infty \end{aligned}$$

so $\alpha f \in \mathrm{BMO}(\mathbb{R}^n)$. Now, let $f, g \in \mathrm{BMO}(\mathbb{R}^n)$. Then

$$\begin{aligned} \|f+g\|_{BMO} &= \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) + g(x) - \mathrm{Avg}_Q\, f - \mathrm{Avg}_Q\, g\right| dx \mid Q \text{ is a cube in } \mathbb{R}^n\right\} \\ &\leq \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) - \mathrm{Avg}_Q\, f\right| dx + \frac{1}{|Q|}\int_Q \left|g(x) - \mathrm{Avg}_Q\, g\right| dx \mid Q \text{ is a cube in } \mathbb{R}^n\right\} \\ &\leq \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) - \mathrm{Avg}_Q\, f\right| dx \mid Q \text{ a cube}\right\} + \sup\left\{\frac{1}{|Q|}\int_Q \left|g(x) - \mathrm{Avg}_Q\, g\right| dx \mid Q \text{ a cube}\right\} \\ &= \|f\|_{BMO} + \|g\|_{BMO} < \infty \end{aligned}$$

thus $f + g \in \mathrm{BMO}(\mathbb{R}^n)$. □

Remark 164. We now know that $\mathrm{BMO}(\mathbb{R}^n)$ is a normed linear space, and as we will see in the next chapter, it is actually a Banach space.

Proposition 165. $\|\cdot\|_{BMO}$ *is seminorm, but not a norm because* $\|f\|_{BMO} = \|f + c\|_{BMO}$ *where* $c \in \mathbb{R}$.

Proof. By Proposition 163, $\|\cdot\|_{BMO}$ is a seminorm. For the second part we have

$$\|f+c\|_{BMO} = \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) + c - \mathrm{Avg}_Q(f+c)\right| dx \mid Q \text{ a cube in } \mathbb{R}^n\right\},$$

but

$$\mathrm{Avg}_Q(f+c) - \frac{1}{|Q|}\int_Q f(x) + c\,dx = \frac{1}{|Q|}\int_Q f(x)dx + c\frac{1}{|Q|}\int_Q dx = \mathrm{Avg}_Q\, f + c.$$

Thus

$$\|f+c\|_{BMO} = \sup\left\{\frac{1}{|Q|}\int_Q \left|f(x) + c - \mathrm{Avg}_Q\, f - c)\right| dx \mid Q \text{ a cube in } \mathbb{R}^n\right\} = \|f\|_{BMO}.$$

As a consequence we have that $\|c\|_{BMO} = 0$, when c is a constant. □

As a result of Proposition 165 we define $f \sim g$ if $f - g = c$ for some constant $c \in \mathbb{R}$.

Proposition 166. *The relation $\sim$ is an equivalence relation.*

Proof. Clearly $f \sim f$. If $f - g = c$, then $g - f = -c$. If $f - g = c$ and $g - h = d$, then $f - g + g - h) = f - h = c + d$, so $f \sim h$, and $\sim$ is an equivalence relation. ☐

We will now identify functions by $\sim$ and define

$$\mathrm{BMO}(\mathbb{R}^n) = \left\{ f \mid \|f\|_{BMO} < \infty \text{ and } f \text{ is locally integrable on } \mathbb{R}^n \right\} / \sim .$$

The following eight propositions are Proposition 7.1.2 in [Gra14b], but we break it into smaller pieces, filling in details for all the proofs.

Proposition 167 ([Gra14b] Proposition 3.1.2)**.** *If $\|f\|_{BMO} = 0$, then f is equal to a constant almost everywhere.*

Proof. Suppose $\|f\|_{BMO} = 0$. Then

$$\|f\|_{BMO} = \sup \left\{ \frac{1}{|Q|} \int_Q \left| f(x) - \mathrm{Avg}_Q\, f \right| dx \mid Q \text{ a cube in } \mathbb{R}^n \right\} = 0$$

which implies that

$$\int_Q \left| f(x) - \mathrm{Avg}_Q\, f \right| dx = 0$$

for all cubes Q in $\mathbb{R}^n$. But then this means that $f(x) - \mathrm{Avg}_Q\, f = 0$ almost everywhere, so $f(x) = \mathrm{Avg}_Q(f)$ almost everywhere, which is a constant. ☐

Example. Let $f = c$ be a constant function. Then the average of f over any cube Q is c, so its mean oscillation

$$\frac{1}{|Q|} \int_Q |c - c|\, dx = 0.$$

Thus the mean oscillation of a constant function is 0. This makes sense because it never deviates from its average.

This next theorem tells us that every bounded function also has bounded mean oscillation. This is one of the most important properties of BMO. As we saw previously the dual space of L^q is L^p if $1 < p < \infty$ and $p^{-1} + q^{-1} = 1$. However, this doesn't work if $p = 1$ or $p = \infty$, the L^p spaces don't behave as well in these settings, so instead we can replace L^∞ with BMO and as we will see in the next chapter we can replace L^1 with H^1 to achieve better duality results.

Proposition 168 ([Gra14b] Proposition 3.1.2)**.** $L^\infty(\mathbb{R}^n) \subset \mathrm{BMO}(\mathbb{R}^n)$ *and* $\|f\|_{BMO} \leq 2\,\|f\|_\infty$.

Proof. Suppose $f \in L^\infty(\mathbb{R}^n)$, then f is bounded, and thus it must be within a bounded distance of its mean, so $f \in \mathrm{CMO}(\mathbb{R}^n)$. Then note that

$$\sup \left\{ \mathrm{Avg}_Q \left| f - \mathrm{Avg}_Q\, f \right| \mid Q \text{ a cube in } \mathbb{R}^n \right\} = \sup \left\{ \frac{1}{|Q|} \int_Q \left| f - \mathrm{Avg}_Q\, f \right| d\mu \mid Q \text{ a cube} \right\} = \|f\|_{BMO}$$

Then

$$\mathrm{Avg}_Q \left| f - \mathrm{Avg}_Q\, f \right| = \frac{1}{|Q|} \int_Q \left| f - \mathrm{Avg}_Q\, f \right| d\mu \leq \frac{1}{|Q|} \int_Q |f(x)|\, d\mu + \frac{1}{|Q|} \left| \mathrm{Avg}_Q\, f \right| d\mu = \mathrm{Avg}_Q\, |f| + \left| \mathrm{Avg}_Q\, f \right|.$$

But then

$$\left|\mathrm{Avg}_Q f\right| = \left|\frac{1}{|Q|}\int_Q f\,d\mu\right| \le \frac{1}{|Q|}\int_Q |f(x)|\,d\mu = \mathrm{Avg}_Q |f|.$$

Giving us

$$\mathrm{Avg}_Q \left|f - \mathrm{Avg}_Q f\right| \le 2\,\mathrm{Avg}_Q |f|.$$

Finally, the average of a function is less than its essential supremum, and giving us

$$2\,\mathrm{Avg}_Q |f| \le 2\,\|f\|_{L^\infty}$$

and since this is true for any cube Q in $\mathbb{R}^n$, it is also true for the supremum, thus

$$\|f\|_{BMO} \le \|f\|^{L^\infty}.$$

□

This next proposition tells us that BMO is closed under taking absolute values, minimums and maximums.

Proposition 169 ([Gra14b] Proposition 3.1.2)**.** *If* $f \in \mathrm{BMO}(\mathbb{R}^n)$*, then so is* $|f|$*. Similarly, if* f, g *are real-valued* BMO *functions, then so are* $\max(f,g)$ *and* $\min(f,g)$*. In other words,* BMO *is a lattice. Moreover,*

$$\begin{aligned}
\||f|\|_{BMO} &\le 2\,\|f\|_{BMO}\\
\|\max(f,g)\|_{BMO} &\le \frac{3}{2}\left(\|f\|_{BMO} + \|g\|_{BMO}\right)\\
\|\min(f,g)\|_{BMO} &\le \frac{3}{2}\left(\|f\|_{BMO} + \|g\|_{BMO}\right).
\end{aligned}$$

Proof. The first statement is a consequence of the fact that

$$\begin{aligned}
\left||f(x)| - \mathrm{Avg}_Q |f|\right| &= \left||f| - \frac{1}{|Q|}\int_Q f(t)\,d\mu\right|\\
&= \left|\frac{1}{|Q|}\int_Q ||f(x)| - f(t)|\,dt\right|\\
&\le \left|\frac{1}{|Q|}\int_Q \left|f(x) - \mathrm{Avg}_Q f + \mathrm{Avg}_Q f - f(t)\right|\,dt\right|\\
&\le \left|\frac{1}{|Q|}\int_Q \left|f(x) - \mathrm{Avg}_Q f\right|\,dt + \frac{1}{|Q|}\int_Q \left|\mathrm{Avg}_Q f - f(t)\right|\,dt\right|\\
&= \left|f - \mathrm{Avg}_Q f\right| + \mathrm{Avg}_Q \left|\mathrm{Avg}_Q f - f\right|.
\end{aligned}$$

Then averaging over Q and taking supremums we see that the right hand side is bounded, and thus the left hand side is as well, so $|f| \in \mathrm{BMO}(\mathbb{R}^n)$. Note that we can write

$$\max(f,g) = \frac{|f-g| + f + g}{2}.$$

Indeed, $\frac{|f-g|}{2}$ is half of the distance between f and g, while $\frac{f+g}{2}$ is the midpoint, so adding half the distance to the midpoint brings us to the maximum. Similarly,

$$\min(f,g) = \frac{f+g-|f-g|}{2}.$$

Now, we have

$$\|\max(f,g)\|_{BMO} = \sup\left\{\frac{1}{|Q|}\int_Q \left|\max(f,g)(x) - \mathrm{Avg}_Q \max(f,g)\right| d\mu \mid Q \text{ is a cube in } \mathbb{R}^n\right\},$$

and

$$\begin{aligned}\mathrm{Avg}_Q(\max(f,g)) &= \frac{1}{|Q|}\int_Q \frac{|f-g|+f+g}{2} d\mu \\ &= \frac{1}{2|Q|}\left(\int_Q |f(x)-g(x)|\, d\mu + \int_Q f(x) d\mu + \int_Q g(x) d\mu\right) \\ &= \mathrm{Avg}_Q |f-g| + \mathrm{Avg}_Q f + \mathrm{Avg}_Q g.\end{aligned}$$

So we have

$$\begin{aligned}\|\max(f,g)\|_{BMO} &= \frac{1}{2}\sup\left\{\frac{1}{Q}\int_Q \left|\frac{|f-g|+f+g}{2} - \mathrm{Avg}_Q |f-g| - \mathrm{Avg}_Q f - \mathrm{Avg}_Q g\right| d\mu \mid Q \text{ a cube in } \mathbb{R}^n\right\} \\ &\leq \left(\||f-g|\|_{BMO} + \|f\|_{BMO} + \|g\|_{BMO}\right).\end{aligned}$$

Using the first result of this proposition, we see that

$$\begin{aligned}\frac{1}{2}\left(\||f-g|\|_{BMO} + \|f\|_{BMO} + \|g\|_{BMO}\right) &\leq \frac{1}{2}\left(2\|f-g\|_{BMO} + \|f\|_{BMO} + \|g\|_{BMO})\right) \\ &\leq \frac{3}{2}\left(\|f\|_{BMO} + \|g\|_{BMO}\right).\end{aligned}$$

The last statement simply follows from the fact that $\min(f,g) \leq \max(f,g)$. □

Proposition 170 ([Gra14b] Example 3.1.3). $L^\infty(\mathbb{R}^n)$ *is a proper subspace of* $BMO(\mathbb{R}^n)$.

Proof. We claim that $\log|x| \in BMO(\mathbb{R}^n)$, but not in $L^\infty(\mathbb{R}^n)$. Clearly $\log|x| \notin L^\infty(\mathbb{R}^n)$ since it is unbounded. To prove that it is in $BMO(\mathbb{R}^n)$, for every $x_0 \in \mathbb{R}^n$ and $R > 0$, we must find a constant $c_{x_0,R}$ such that the average of $\left|\log|x| - c_{x_0,R}\right|$ over the ball $\{x \in \mathbb{R}^n \mid |x - x_0| \leq R\}$ is uniformly bounded. Since

$$\frac{1}{v_n R^n}\int_{|x-x_0|\leq R} \left|\log|x| - c_{x_0,R}\right| d\mu = \frac{1}{v_n}\int_{|z-R_{-1}x_0|\leq 1} \left|\log|z| - c_{x_0,R} + \log R\right| d\mu,$$

we may take $c_{x_0,R} = c_{R^{-1}x_0,1} + \log R$, and things reduce to the case that $R = 1$ and x_0 is arbitrary. If $R = 1$ and $|x_0| \leq 2$, take $c_{x_0,1} = 0$ and observe that

$$\int_{|x-x_0|\leq 1} |\log|x||\, d\mu \leq \int_{|x|\leq 3} |\log|x||\, d\mu = c.$$

When $R = 1$ and $|x_0| \geq 2$, take $c_{x_0,1} = \log|x_0|$. In this case notice that

$$\frac{1}{v_n}\int_{|x-x_0|\leq 1} |\log|x| = \log|x_0||\, dx = \frac{1}{v_n}\int_{|x-x_0|\leq 1} \left|\log\frac{|x|}{|x_0|}\right| \leq \log 2.$$

Since when $|x - x_0| \leq 1$ and $|x_0| \geq 2$, we have

$$\log\frac{|x|}{|x_0|} \leq \log\frac{|x_0|+1}{|x_0|} \leq \log\frac{3}{2}$$

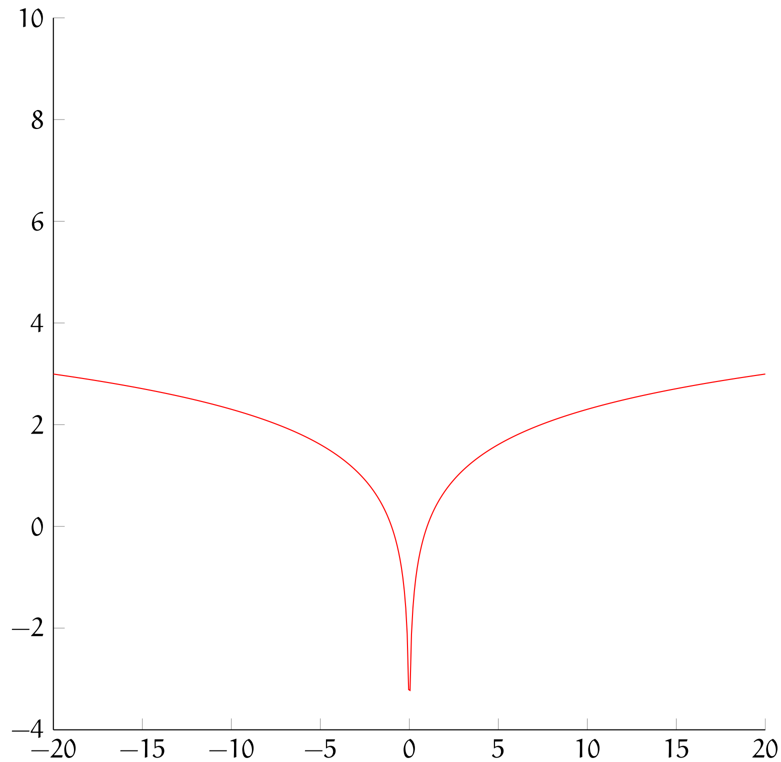


Figure 16.1: Graph of $\log|x|$

which comes from the assumption that $|x_0| \geq 2$, and

$$\log\frac{|x_0|}{|x|} \leq \log\frac{|x_0|}{|x_0|-1} \leq \log 2.$$

Thus $\log x \in \mathrm{BMO}(\mathbb{R}^n)$. □

Example ([Gra14b] Example 3.1.4)**.** Now we show that not all functions are in $\mathrm{BMO}(\mathbb{R}^n)$. We claim that $h(x) = \chi_{x>0}\log\frac{1}{x} \notin \mathrm{BMO}(\mathbb{R}^n)$, where $\chi_{x>0}$ is the characteristic function. The problem is at the origin. Consider the intervals $(-\epsilon,\epsilon)$ with $0<\epsilon<\frac{1}{2}$. We have that

$$\mathrm{Avg}_{(-\epsilon,\epsilon)}h = \frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon} h(x)dx = \frac{1}{2\epsilon}\int_0^{\epsilon}\log\frac{1}{x}dx = \frac{1+\log\frac{1}{\epsilon}}{2}.$$

But then

$$\frac{1}{2\epsilon}\int_{-\epsilon}^{\epsilon}\left|h(x)-\mathrm{Avg}_{(-\epsilon,\epsilon)}h\right|d\mu \geq \frac{1}{2\epsilon}\int_{-\epsilon}^{0}\left|\mathrm{Avg}_{(-\epsilon,\epsilon)}h\right|d\mu = \frac{1+\log\frac{1}{\epsilon}}{4},$$

but the right hand side of this is unbounded as $\epsilon\to 0$.

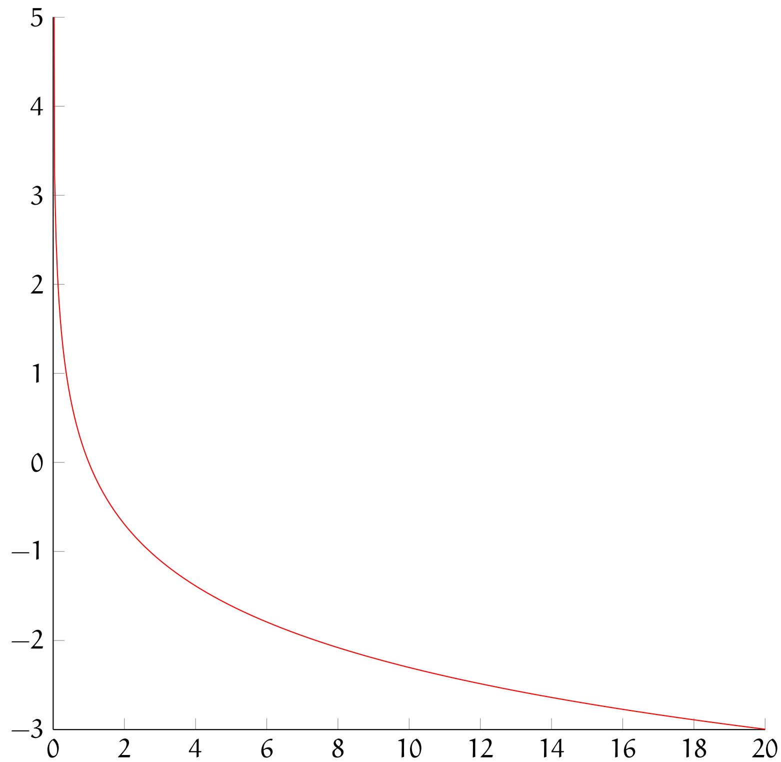


Figure 16.2: Graph of $\log \frac{1}{x}$

16.1 Calderón-Zygmund Decomposition

We want to prove the John-Nirenberg inequality which says that BMO functions are exponentially integrable. However, we first need to make a detour to understand the Calderón Zygmund Decomposition.

Definition 171. A *dyadic cube* in $\mathbb{R}^n$ is the set

$$[2^k m_1, 2^k(m_1+1)) \times \cdots \times [2^k m_n, 2^k(m_n+1)),$$

where $k, m_1, \ldots, m_n \in \mathbb{Z}$. Two dyadic cubes are either disjoint or related by inclusion.

Example. Put an example in here.

We also need one more theorem, which we state without proof.

Theorem 172 (Lebesgue's Differentiation Theorem, [Gra14a] Corollary 2.1.16)**.** *For any locally integrable function f on $\mathbb{R}^n$ we have*

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

for almost all $x \in \mathbb{R}^n$.

Remark 173. Lebesgue's differentiation theorem says that the value of a function at a point is the limit of its average on the ball surrounding it.

Now we come to the Calderón-Zygmund decomposition which says that for any absolutely integrable function we can decompose it into a good part and a bad part. The bad part also has mean zero and is also in L^1.

Theorem 174 ([Gra14a] Corollary 5.3.1)**.** *Let $f \in L^1(\mathbb{R}^n)$ and $\alpha > 0$. Then there exist functions g and b on $\mathbb{R}^n$ such that*

(a) $f = g + b$.

(b) $\|g\|_{L^1} \leq \|f\|_{L^1}$ *and* $\|g\|_{L^\infty} \leq 2^n\alpha$.

(c) $b = \sum_j b_j$ *where each* b_j *is supported in a dyadic cube* Q_j. *Furthermore, the cubes* Q_k *and* Q_j *are disjoint when* $j \neq k$.

(d) $\int_{Q_j} b_k(x)dx = 0$.

(e) $\|b_j\|_{L^1} \leq 2^{n+1}\alpha\left|Q_j\right|$.

(f) $\sum_j \left|Q_j\right| \leq \alpha^{-1}\|f\|_{L^1}$.

Proof. We first construct the cubes Q_j. Start by decomposing $\mathbb{R}^n$ into a mesh of disjoint dyadic cubes of the same size such that

$$|Q| \geq \frac{1}{\alpha}\|f\|_{L^1}$$

for every cube Q in the mesh. We call these cubes generation zero. To build the next generation, subdivide each cube of generation zero in 2^n congruent cubes by bisecting each of its sides. This will be a new mesh of dyadic cubes of generation one. Now, choose a cube Q of generation one if

$$\frac{1}{|Q|}\int_Q |f(x)|\,dx > \alpha.$$

We are singling out the cubes where f is too big, and these will be the cubes we define the b_j on. Let $S^{(1)}$ be the set of all cubes selected in generation one. Now subdivide each nonselected cube of generation one into 2^n congruent subcubes by dissecting the sides. These cubes will be generation two. Then let $S^{(2)}$ be the set of cubes such that

$$\frac{1}{|Q|}\int_Q |f(x)|\,dx > \alpha.$$

Repeat this procedure indefinitely. The set of all selected cubes $\bigcup_{m=1}^{\infty} S^{(M)}$ is countable and is exactly the set of cubes Q_j that we want for the proposition. Indeed, for each selected sube we have

$$\begin{aligned}\frac{1}{|Q|}\int_Q |f(x)|\,dx &> \alpha.\\ \frac{1}{\alpha}\int_Q |f(x)|\,dx &> \left|Q_j\right|\\ \frac{1}{\alpha}\|f\|_{L^1} &> \sum_j \left|Q_j\right|.\end{aligned}$$

Note that the cubes Q_j are disjoint, since otherwise some Q_k would be a proper subset of some Q_j, which is impossible since the selected cube Q_j was never subdivided. Now define

$$b_j\left(f-\frac{1}{\left|Q_j\right|}\int_{Q_j} f dx\right)\chi_{Q_j},$$

$b=\sum_j b_j$ and $g = f - b$. For a selected cube Q_j there exists a unique nonselected cube Q' with twice its side length that contains Q_j. Call Q' the parent of Q_j. Since the parent Q' of Q_j was not selected, we have $\frac{1}{|Q'|}\int_{Q'}|f|\,dx \le \alpha$. Then

$$\frac{1}{\left|Q_j\right|}\int_{Q_j}.\,|f(x)|\,dx \le \frac{1}{\left|Q_j\right|}\int_{Q'}|f(x)|\,dx = \frac{2^n}{|Q'|}\int_{Q'}|f(x)|\,dx \le 2^n\alpha.$$

And thus

$$\int_{Q_j}\left|b_j\right|dx \le \int_{Q_j}|f|\,dx + \left|Q_j\right|\left|\frac{1}{\left|Q_j\right|}\int_{Q_j} f dx\right| \le 2\int_{Q_j}|f|\,dx \le 2^{n+1}\alpha\left|Q_j\right|.$$

Now we need to obtained the estimates for g. We have

$$g=\begin{cases} f & \text{on } \mathbb{R}^n\setminus\bigcup_j Q_j,\\ \frac{1}{|Q_j|}\int_{Q_j} f dx & \text{on } Q_j.\end{cases}$$

On the cube Q_j, g is equal to the constant $\left|Q_j\right|^{-1}\int_{Q_j} f dx$, which is bounded by $2^n\alpha$. It suffices to show that g is bounded outside the union of the Q_j's. Indeed, for each $x \in \mathbb{R}^n\setminus\bigcup_j Q_j$ and for each $k = 0, 1, 2, \dots$ there exists a unique nonselected dyadic cube $Q_x^{(k)}$ of generation k that contains x. Then for each $k \ge 0$ we have

$$\left|\frac{1}{\left|Q_x^{(k)}\right|}\int_{Q_x^{(k)}} f(y) dx\right| \le \frac{1}{\left|q_x^{(k)}\right|}\int_{Q_x^{(k)}}|f(y)|\,dy \le \alpha.$$

The intersection of the closures of the cubes $Q_x^{(k)}$ is a singleton $\{x\}$. Using Lebesgue's Differentiation Theorem, we have that for almost all $x \in \mathbb{R}^n \setminus \bigcup_j Q_j$ we have

$$f(x) = \lim_{k\to\infty} \frac{1}{\left|Q_x^{(k)}\right|} \int_{Q_x^{(k)}} f(y)dy.$$

Since these averages are at most α, we conclude that $|f| \le \alpha$ almost everywhere on $\mathbb{R}^n \setminus \bigcup_j Q_j$, and hence $|g| \le \alpha$ almost everywhere on this set. Finally, it follows that $\|g\|_{L^1} \le \|f\|_{L^1}$. □

16.2 John-Nirenberg Inequality

We now come to one of the most important inequalities in the theory of BMO spaces. The proof is long, and we have tried to fill in as many details as possible. This section is based off of section 7.1.2 in [Gra14b].

Theorem 175 ([Gra14b] Corollary 3.1.6)**.** *For all* $f \in \mathrm{BMO}(\mathbb{R}^n)$*, for all cubes* Q*, and all* $\alpha > 0$ *we have*

$$\left|\left\{x \in Q\colon \left|f(x) - \mathrm{Avg}_Q f\right| > \alpha\right\}\right| \le e^{-A\alpha/\|f\|_{BMO}} \tag{16.1}$$

with $A = (2^n e)^{-1}$.

Proof. Since the inequality in 16.1 is unchanged when we multiply both f and α by a constant, we can assume that $\|f\|_{BMO} = 1$. We now fix a closed cube Q and a constant $b > 1$ which we will assign a value to later.

We apply Calderón Zygmund decomposition to the function $f - \mathrm{Avg}_Q f$ inside the cube Q. The decomposition will differ slightly since in the Calderón-Zygmund decomposition we looked at before, we were decomposing a function on $\mathbb{R}^n$, but the proof will be similar. We introduce the following selection criterion for a cube R:

$$\frac{1}{|R|}\int_R \left|f(x) - \mathrm{Avg}_Q f\right| dx > b.$$

Since

$$\frac{1}{|Q|}\int_Q \left|f(x) - \mathrm{Avg}_Q f\right| dx \le \|f\|_{BMO} = 1 < b$$

Q will not be chosen. Set $Q^{(0)} = Q$ and subdivide $Q^{(0)}$ into 2^n equal subcubes of side length equal to half of the side length of Q. Select a subcube R if it satisfies the selection criterion. Now subdivide all nonselected cubes into 2^n equal subcubes of half their sidelength by bisecting the sides, and select the cubes that meet the selection criterion. By continuing this process indefinitely, we will obtain a countable collection of subcubes $\left\{Q_j^{(1)}\right\}_j$ satisfying the following properties:

(A-1) The interior of every $Q_j^{(1)}$ is contained in $Q^{(0)}$.

(B-1) $b < \left|Q_j^{(1)}\right|^{-1} \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}} f\right| dx \le 2^n b$.

(C-1) $\left|\mathrm{Avg}_{Q_j^{(1)}} f - \mathrm{Avg}_{Q^{(0)}} f\right| \le 2^n b$.

(D-1) $\sum_j \left|Q_j^{(0)}\right| \le \frac{1}{b}\sum_j \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx \le \frac{1}{b}\left|Q^{(0)}\right|$.

(E-1) $\left|f - \mathrm{Avg}_{Q^{(0)}} f\right| \leq b$ almost everywhere on the set $Q^{(0)} \setminus \bigcup_j Q_j^{(1)}$.

We prove these quickly. Properties (A-1) and the first inequality follow from construction. For the second inequality in (B-1) suppose that

$$2^n b < \frac{1}{\left|Q_j^{(1)}\right|} \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx,$$

then

$$b < \frac{1}{2^n \left|Q_j^{(1)}\right|} \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx \leq \frac{1}{2^n \left|Q_j^{(1)}\right|} \int_{2^n Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx$$

but $2^n Q_j^{(1)}$ is the size of the cube containing $Q_j^{(1)}$ which was not selected. Thus by this it should have been selected which is a contradiction, establishing (B-1). For (C-1) we have

$$\begin{aligned}
\left|\mathrm{Avg}_{Q_j^{(1)}} f - Q^{(0)}\right| &= \left|\frac{1}{\left|Q_j^{(1)}\right|} \int_{Q_j^{(1)}} f(x)dx - \mathrm{Avg}_{Q^{(0)}} f\right| \\
&= \left|\frac{1}{\left|Q_j^{(1)}\right|} \int_{Q_j^{(1)}} f(x) - \mathrm{Avg}_{Q^{(0)}} f dx\right| \\
&\leq \frac{1}{\left|Q_j^{(1)}\right|} \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx \leq 2^n b
\end{aligned}$$

by the second inequality in (B-1). For (D-1) the first inequality just follows from the first inequality in (B-1). For the second inequality note that

$$\frac{1}{\left|Q^{(0)}\right|} \int_{Q^{(0)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx \leq 1 < b.$$

Then

$$\frac{1}{b} \int_{Q^{(0)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right| dx \leq \frac{1}{b} \left|Q^{(0)}\right|.$$

Note that

$$\int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}} f\right| dx \leq \int_{Q^{(0)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}} f\right| dx,$$

in fact,

$$\sum_j \int_{Q_j^{(1)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}} f\right| dx \leq \int_{Q^{(0)}} \left|f(x) - \mathrm{Avg}_{Q^{(0)}} f\right| dx,$$

since the $Q_j^{(1)}$ are all disjoint and contained inside $Q^{(0)}$. Thus we get the second inequality in (D-1). Finally, for (E-1) we can apply the Lebesgue differentiation theorem to the function $\left|f(x) - \mathrm{Avg}_{Q^{(0)}}\right|$ to obtain the result.

We call the cubes $Q_j^{(1)}$ the first generation. We now fix a selected first generation cube $Q_j^{(1)}$ and we introduce the following selection criteria for a cube R:

$$\frac{1}{|R|} \int_R \left|f(x) - \mathrm{Avg}_{Q_j^{(1)}} f\right| dx > b.$$

Similar to the decomposition we did before, $Q_j^{(1)}$ does not satisfy the selection criteria. We again apply a Calderón-Zygmund decomposition to the function $f - \operatorname{Avg}_{Q_j^{(1)}} f$ inside the cube $Q_j^{(1)}$. Subdivide $Q_j^{(1)}$ into 2^n equal closed subcubes of $Q_j^{(1)}$. Select a subcube R if it satisfies the selection criteria. Then divide the nonselected cubes into 2^n subcubes and repeat. Continue this process indefinitely. Also, repeat this process for any other cube $Q_j^{(1)}$ in the first generation. We obtain a collection of cubes $\left\{Q_l^{(2)}\right\}_l$ that are the second generation cubes. Versions of (A-1)-(E-1) are satisfied by replacing the superscript (1) with (2) and with (1) replacing (0). We use the superscript (k) to denote the generation of the selected cubes.

For a fixed selected cube $Q_l^{(2)}$ in the second generation we introduce the selection criteria

$$\frac{1}{|R|}\int_R \left|f(x) - \operatorname{Avg}_{Q_l^{(2)}} f\right| dx > b$$

and repeat the previous process to obtain a third generation of cubes. Denote by $\left\{Q_s^{(3)}\right\}_s$ this third generation. We iterate this procedure indefinitely to obtain a doubly indexed family of cubes $Q_j^{(k)}$ satisfying the following properties.

(A-k) The interior of every $Q_j^{(k)}$ is contained in a unique $Q_{j'}^{(k-1)}$.

(B-k) $b < \left|Q_j^{(k)}\right|^{-1} \int_{Q_j^{(k)}} \left|f(x) - \operatorname{Avg}_{Q_{j'}^{(k-1)}} f\right| dx \leq 2^n b$.

(C-k) $\left|\operatorname{Avg}_{Q_j^{(k)}} f - \operatorname{Avg}_{Q_{j'}^{(k-1)}} f\right| \leq 2^n b$.

(D-k) $\sum_j \left|Q_j^{(k)}\right| \leq \frac{1}{b} \sum_{j'} \left|Q_{j'}^{(k-1)}\right|$.

(E-k) $\left|f - \operatorname{Avg}_{Q_{j'}^{(k-1)}} f\right| \leq b$ almost everywhere on the set $Q_{j'}^{(k-1)} \setminus \bigcup_j Q_j^{(k)}$.

Note that (A-1) and the first inequality of (B-1) are satisfied by construction. Then similar to the last case the second inequality in (B-1) is a consequence of the fact that the unique cube $Q_{j_0}^{(k)}$ with double the side length of $Q_j^{(k)}$ that contains it was not selected in the process. Then (C-K) follows from the upper inequality in (B-K). Indeed,

$$\begin{aligned}\left|\operatorname{Avg}_{Q_j^{(k)}} - Q_{j'}^{(k-1)}\right| &= \left|\frac{1}{\left|Q_j^{(k)}\right|}\int_{Q_j^{(k)}} \left|f(x) - \operatorname{Avg}_{Q_{j'}^{(k-1)}} f\right| dx\right| \\ &\leq \frac{1}{\left|Q_j^{(k)}\right|}\int_{Q_j^{(k)}} \left|f(x) - \operatorname{Avg}_{Q_{j'}^{(k-1)}} f\right| dx \leq 2^n b.\end{aligned}$$

Next, (E-k) follows from the Lebesgue Differentiation theorem in the same way that (E-1) did. Finally, for (D-k) we have

$$\begin{aligned}\sum_j \left|Q_j^{(k)}\right| &\leq \frac{1}{b}\sum_j \int_{Q_j^{(k)}} \left|f(x) - \mathrm{Avg}_{Q_{j'}^{(k-1)}} f\right| dx \\ &= \frac{1}{b}\sum_{j'} \sum_{j \text{ corresp. to } j'} \int_{Q_j^{(k)}} \left|f(x) - \mathrm{Avg}_{Q_{j'}^{(k-1)}} f\right| dx \\ &\leq; \sum_{j'} \int_{Q_{j'}^{(k-1)}} \left|f(x) - \mathrm{Avg}_{Q_{j'}^{(k-1)}} f\right| dx \\ &\leq \frac{1}{b}\sum_{j'} \left|Q_{j'}^{(k-1)}\right| \|f\|_{BMO} \\ &= \frac{1}{b}\sum_{j'} \left|Q_{j'}^{(k-1)}\right|.\end{aligned}$$

Now we look at some consequences of (A-k)-(E-K). If we apply (D-k) $k-1$ times successively we get

$$\sum_j \left|Q_j^{(k)}\right| \leq b^{-k} \left|Q^{(0)}\right|.$$

For any fixed j we have that

$$\left|\mathrm{Avg}_{Q_j^{(1)}} f - \mathrm{Avg}_{q^{(0)}}\right| \leq 2^n b$$

and $\left|f - \mathrm{Avg}_{Q_j^{(1)}} f\right| \leq b$ almost everywhere on $Q_j^{(1)} \setminus \bigcup_l Q_l^{(2)}$. This gives

$$\left|f - \mathrm{Avg}_{Q^{(0)}} f\right| \leq 2^n b + b$$

almost everywhere on $Q_j^{(1)} \setminus \bigcup_l Q_l^{(2)}$. Indeed,

$$\left|f - \mathrm{Avg}_{Q^{(0)}} f\right| = \left|f - \mathrm{Avg}_{Q_j^{(1)}} f + \mathrm{Avg}_{Q_j^{(1)}} - \mathrm{Avg}_{Q^{(0)}} f\right| \leq \left|f - \mathrm{Avg}_{Q_j^{(1)}} f\right| + \left|\mathrm{Avg}_{Q_j^{(1)}} - \mathrm{Avg}_{Q^{(0)}} f\right| \leq 2^n b + b$$

almost everywhere on $Q_j^{(1)} \setminus \bigcup_l Q_l^{(2)}$. We will use the fact that $2^n b + b \leq 2^n 2b$ combined with (E-1) to see that

$$\left|f - \mathrm{Avg}_{Q^{(0)}}\right| \leq 2^n 2b$$

almost everywhere on $Q^{(0)} \setminus \bigcup_l Q_l^{(2)}$. Now, for every fixed l we also have that $\left|f - Avg_{Q_l^{(2)}} f\right| \leq b$ almost everywhere on $Q_l^{(2)} \setminus \bigcup_s Q_s^{(3)}$, which combined with $\left|\mathrm{Avg}_{Q_l^{(2)}} f - \mathrm{Avg}_{Q_{l'}^{(1)}} f\right| \leq 2^b$ and $\left|\mathrm{Avg}_{Q_{l'}^{(1)}} f - \mathrm{Avg}_{Q^{(0)}} f\right| \leq 2^n b$ yields

$$\left|f - \mathrm{Avg}_{Q^{(0)}} f\right| \leq 2^n 3b$$

almost everywhere on $Q_l^{(2)} \setminus \bigcup_s Q_s^{(3)}$. And the same estimate is valid on $Q^{(0)} \setminus \bigcup_s Q_s^{(3)}$ Continuing this way by induction we have that for all $k \geq 1$,

$$\left|f - \mathrm{Avg}_{Q^{(0)}} f\right| \leq 2^n kb$$

almost everywhere on $Q^{(0)} \setminus \bigcup_s Q_s^{(k)}$. This proves the following inclusion almost everywhere

$$\left\{ x \in Q \mid \left| f(x) - \mathrm{Avg}_Q f \right| > 2^n kb \right\} \subseteq \bigcup_j Q_j^{(k)}$$

for all $k \in \mathbb{N}$. Now, fix $\alpha > 0$. If

$$2^n kb < \alpha \leq 2^n (k+1) b$$

for some $k \geq 0$, then

$$\begin{aligned} \left| \left\{ x \in Q \mid \left| f - \mathrm{Avg}_Q f \right| > \alpha \right\} \right| &\leq \left| \left\{ x \in Q \mid \left| f - \mathrm{Avg}_Q f \right| > 2^n kb \right\} \right| \\ &\leq \sum_j \left| Q_j^{(k)} \right| \leq \frac{1}{b^k} \left| Q \right| \\ &= \left| Q \right| e^{-k \log(b)} \\ &\leq \left| Q \right| b e^{-\alpha \log(b)/(2^n b)} \end{aligned}$$

where the last inequality comes from the fact that $-k \leq 1 - \frac{\alpha}{2^n b}$. Choosing $b = e > 1$ yields the inequality. □

Chapter 17

BMO on Shapes

Now we want to look at a variant of BMO spaces. In the definition of the $BMO(\mathbb{R}^n)$ norm we take the supremum over all cubes of the mean oscillation, however, there is nothing special about cubes. We can define BMO on any set of shapes. These spaces are called BMO on shapes and were first introduced by Dafni and Gibara in [DG20]. There are a couple of minimal requirements we need this set of shapes to satisfy.

Definition 176 ([DG20] Definition 2.1)**.** A *shape* in $\mathbb{R}^n$ is any open set in S such that $0 < |S| < \infty$. We call a collection $\mathcal{S}$ of shapes a *basis* if each $S \in \mathcal{S}$ is a shape and $\bigcup_{S \in \mathcal{S}} S = \mathbb{R}^n$.

Example. In one dimension all shapes are the same, as they are just intervals. In higher dimensions there are many differences between different bases of shapes. For example, let $\mathcal{R}$ be the basis of rectangles with sides parallel to the axis, and let $\mathcal{Q}$ be the basis of cubes with sides parallel to the axis. Then these bases are very different because a sequence of rectangles can have their measure go to zero while the diameter does not. This does not happen in $\mathcal{Q}$. Other examples included balls centered at the origin or rectangles without sides parallel to the axes. See some more examples in figure 17.1.

From this we can then define BMO on shapes.

Definition 177 ([DG20] Definition 3.1)**.** Let $\mathcal{S}$ be a basis of shapes on $\mathbb{R}^n$. A function f such that $f \in L^1(S)$ for all

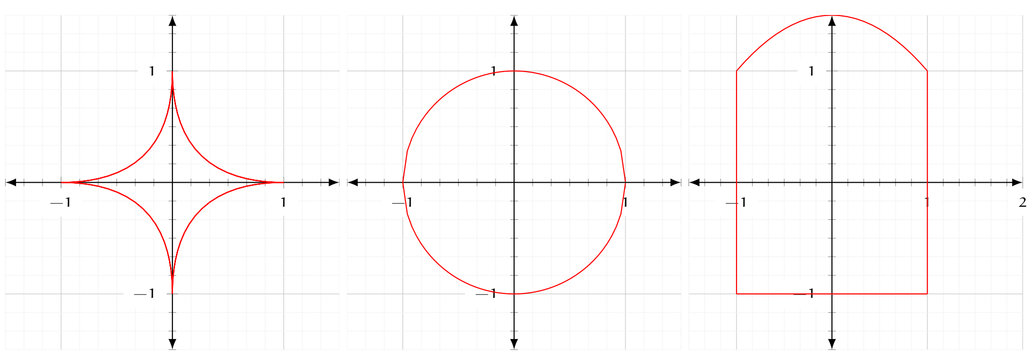


Figure 17.1: Examples of Shapes in $\mathbb{R}^2$

$S \in \mathcal{S}$ is in the space $BMO_{\mathcal{S}}(\mathbb{R}^n)$ if

$$\sup_{S\in\mathcal{S}} \frac{1}{|S|}\int_S |f - f_S|\,dx$$

is bounded, where $f_S = \frac{1}{|S|}\int_S f\,dx$.

Before we move onto the main duality result we prove that $BMO_{\mathcal{S}}(\mathbb{R}^n)$ is indeed a Banach space. We follow the proof of [DG20] which does not pass to the dual space as the classical completeness proofs for $BMO_{\mathcal{Q}}(\mathbb{R}^n)$ do.

Theorem 178 ([DG20] Theorem 3.9)**.** *For any basis of shapes,* $BMO_{\mathcal{S}}(\mathbb{R}^n)$ *is complete.*

Proof. Let $\{f_i\}$ be a Cauchy sequence in $BMO_{\mathcal{S}}(\mathbb{R}^n)$. Then, for any shape $S \in \mathcal{S}$, the sequence $\{f_i - (f_i)_S\}$ is Cauchy in $L^1(S)$. Since $L^1(S)$ is complete, there exists a function $f^S \in L^1(S)$ such that $f_i - (f_i)_S \to f^S$ in $L^1(S)$. Note that f^S has mean zero on S. Indeed, since $f_i - (f_i)_S$ converges to f^S in $L^1(S)$, it follows that

$$\frac{1}{|S|}\int_S f^S = \lim_{i\to\infty}\frac{1}{|S|}\int_S f_i - (f_i)_S = 0.$$

If we have two shapes $S_1, S_2 \in \mathcal{S}$ such that $S_1 \cap S_2 \neq \emptyset$, by the above there is a function $f^{S_1} \in L^1(S_1)$ such that $f_i - (f_i)_{S_1} \to f^{S_1}$ in $L^1(S_1)$ and a function $f^{S_2} \in L^1(S_2)$ such that $f_i - (f_i)_{S_2} \to f^{S_2}$ in $L^1(S_2)$. Since both of these hold in $L^1(S_1 \cap S_2)$, we have

$$(f_i)_{S_2} - (f_i)_{S_1} = [f_i - (f_i)_{S_1}] - [f_i - (f_i)_{S_2}] \to f^{S_1} - f^{S^2}$$

int $L^p(S_1 \cap S_2)$. This implies hat the sequence $C(S_1, S_2) = (f_i)_{S_2} - (f_i)_{S_1}$ converges as constants to a limit that we denote by $C(S_1, S_2)$, with

$$f^{S_1} - f^{S_2} = C(S_1, S_2) \text{ on } S_1 \cap S_2.$$

Note that these constants are antisymmetric, i.e., $C(S_1, S_2) = -C(S_2, S_1)$.

By a *finite chain* of shapes we mean a finite sequence $\{S_j\}_{j=1}^k \subset \mathcal{S}$ such that $S_j \cap S_{j+1} \neq \emptyset$ for all $1 \leq j \leq k-1$. Furthermore, by a *loop* of shapes we mean a finite chain $\{S_j\}_{j=1}^k$ such that $S_1 \cap S_k \neq \emptyset$. If $\{S_j\}_{j=1}^k$ is a loop of shapes, then

$$C(S_1, S_k) = \sum_{j=1}^{k-1} C(S_j, S_{j+1}).$$

To see this, consider the telescoping sum

$$(f_i)_{S_k} - (f_i)_{S_1} = \sum_{j=1}^{k-1} (f_i)_{S_{j+1}} - (f_i)_{S_j},$$

for a fixed i. The above formula follows from this as each $(f_i)_{S_{j+1}} - (f_i)_{S_j}$ converges to $C(S_j, S_{j+1})$ since $S_j \cap S_{j+1} \neq \emptyset$ and $(f_i)_{S_k} - (f_i)_{S_1}$ converges to $C(S_1, S_k)$ since $S_1 \cap S_k \neq \emptyset$.

Now we fix a shape $S_0 \in \mathcal{S}$ and consider another shape $S \in \mathcal{S}$ such that $S_0 \cap S = \emptyset$. For any pair of points $(x, y) \in S_0 \times S$ there exists a path $\gamma_{x,y} \colon [0, 1] \to \mathbb{R}^n$ such that $\gamma_{x,y}(0) = x$ and $\gamma_{x,y}(1) = y$. Since $\mathcal{S}$ covers $\mathbb{R}^n$ and the image of $\gamma_{x,y}$ is a compact set, we may cover $\gamma_{x,y}$ by a finite number of shapes. From this we may extract a finite chain connecting S to S_0.

Now we build the limit function f. If $x \in S_0$, then set $f(x) = f^{S_0}(x)$. If $x \notin S_0$, then there is some shape S containing x and by a preceding argument, a finite chain of shapes $\{S_j\}_{j=1}^k$ where $S_k = S$. In this case, set

$$f(x) = f^{S_k}(x) + \sum_{j=0}^{k-1} C(S_j, S_{j+1}).$$

To see that this is well-defined, let $\{\tilde{S_j}\}_{j=1}^l$ be another finite chain connecting some $\tilde{S}_l$ with $x \in \tilde{S}_l$ to $S_0 = \tilde{S}_0$. Then we need to show that

$$f^{S_k} + \sum_{j=0}^{k-1} C(S_j, S_{j+1}) = f^{\tilde{S}_l} + \sum_{j=0}^{l-1} C(\tilde{S}_j, \tilde{S}_{j+1}).$$

First, we use the fact that $x \in S_k \cap \tilde{S}_l$ to write $f^{S_k} - f^{\tilde{S}_l}(x) = C(S_k, \tilde{S}_l)$. Then, from the antisymmetry property of the constants, this is equivalent to

$$C(S_k, \tilde{S}_l) = C(S_k, S_{k-1}) + \cdots + C(S_1, S_0) + C(S_0, \tilde{S}_1) + \cdots + C(\tilde{S}_{l-1}, \tilde{S}_l).$$

Finally, we show that $f_i \to f$ in $BMO_{\mathcal{S}}(\mathbb{R}^n)$. Fixing a shape $S \in \mathcal{S}$, choose a finite chain $\{S_j\}_{j=1}^k$ such that $S_k = S$. Then we have that, on S $f = F^S$ modulo constants, and so, using the definition of f^S we get

$$\begin{aligned} \frac{1}{|S|}\int_S |(f_i(x) - f(x)) - (f_i - f)_S|\,dx &= \frac{1}{|S|}\int_S \left|(f_i(x) - f^S(x)) - (f_i - f^S)_S\right| dx \\ &= \frac{1}{|S|}\int_S \left|f_i(x) - (f_i)_S - f^S(x)\right| dx \to 0 \end{aligned}$$

as $i \to \infty$. □

17.1 H^1 − BMO **Duality**

One of the most important theorems of harmonic analysis in the second half of the twentieth century was the duality between the Hardy space H^1 and BMO by Charles Fefferman in [Fef71]. The space H^1 has many different characterizations, but we will define it using atomic decomposition. It is usually defined first as a space of holomorphic functions, or more generally, as a space of tempered distributions subject to a maximal condition. It is then possible to show that these are all equivalent but we will not do that here. This is because the equivalency of these definitions relies heavily on the fact that they are defined using cubes, and does not work as well for shapes. However, the atomic characterization easily translates over to shapes. The outline of the rest of this thesis is that we will define a Hardy space on shapes and show that its dual space is the BMO on shapes that we defined above.

Hardy Spaces were originally introduced in the following way.

Definition 179 (Hardy Space)**.** The Hardy space H^p for $0 < p < \infty$ is the class of holomorphic functions f on the open unit disk that satisfy

$$\sup_{0 \le r < 1} \left(\frac{1}{2\pi}\int_0^{2\pi} \left|f(re^{i\theta})\right|^2 d\theta \right)^{1/p} < \infty.$$

Recall that for a fixed r and $0 \le \theta \le 2\pi$, the quantity $re^{i\theta}$ is the circle centered at the origin of radius r. So we are almost taking the average value of the function over the unit circle. There are a couple of other important things to note about this definition. First, the norm is very similar to the L^p norm, but we allow $0 < p < 1$. Recall that if $p < 1$, then L^p is no longer a normed space. This makes the Hardy spaces a possible stand in for the classical L^p spaces when $p < 1$.

The way that our duality theorem will eventually be proven is by using a different characterization of Hardy spaces, called atomic decomposition. We introduce this below.

The idea of atomic decomposition is to take a function and write it as a linear combination of much simpler functions. We will, in a sense, be reversing this by defining our Hardy space on shapes to be all linear combinations of such functions. We start with the definition of these simple functions, called atoms.

Definition 180 (Atom)**.** Let $\mathcal{S}$ be a basis of shapes and let $S \in \mathcal{S}$. Then a measurable function $a\colon \mathbb{R}^n \to \mathbb{C}$ is called an *atom on shapes* if

(a) $\operatorname{Supp} a \subset S$.

(b) $\|a\|_2 \leq \frac{1}{|S|^{1/2}}$

(c) $\int_S a\,dx = 0$.

Remark 181. A more conventional definition would change item (2) to $|a(x)| \leq |S|^{-1}$ almost everywhere, but these two conditions are equivalent. Indeed, this inequality can be rewritten as

$$|S|^{1/2}\left(\int_S |a|^2\,dx\right)^{1/2} \leq 1.$$

Then, by Hölder's inequality we have

$$\int_S |a|dx \leq |S|^{1/2}\left(\int_S |a|^2\,dx\right)^{1/2} \leq 1.$$

So

$$\int_S |a|dx \leq 1 = \int_S \frac{1}{|S|}$$

and thus $|a| \leq |S|^{-1}$.

Example. Suppose our basis of shapes is the set of all intervals on $\mathbb{R}$. Let $I = [-1, 1]$ and define

$$a(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 1 \\ -\frac{1}{2} & \text{if } -1 \leq x < 0. \end{cases}$$

See figure 17.2.

Now we can define the Hardy space on shapes.

Definition 182. Let

$$H^1_{\mathcal{S}} = \left\{ \sum_{i=1}^{\infty} \lambda_i a_i \mid a_i \text{ is an atom on shapes and } \lambda_i \in \mathbb{C}, \sum_{i=1}^{\infty} |\lambda_i| < \infty \right\}.$$

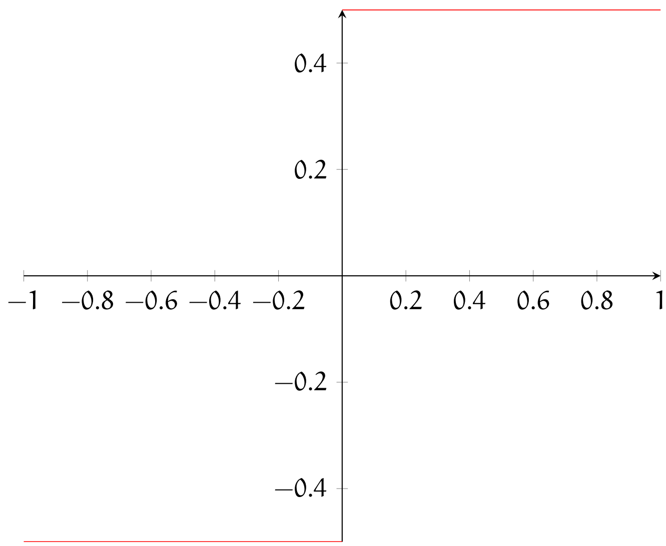


Figure 17.2: Example of an atom

Remark 183. In the set H^1_S there may be some functions appearing twice since there could be multiple ways to represent a typical function as a linear combination of atoms. Also, note that H^1_S is a linear space since adding two linear combinations of atoms and multiplying by a scalar is clearly still in the space.

Lastly, we need to put a norm on this set to show that it is a normed linear space. Let $f \in H^1_S$ and define

$$\|f\|_{H^1_S} = \inf\left\{\sum_{i=1}^{\infty} |\lambda_i| \mid f = \sum_{i=1}^{\infty} \lambda_i a_i\right\}$$

The idea is that since there may be many different ways of representing one particular function, we want to take the "smallest" way to and define that to be the norm.

The last important idea we will need for the proof is that of a dense linear subspace. When we have infinite series sometimes the integral of such a function will not converge absolutely. So instead we will take the following subspace

$$H^1_{S,a} = H^1_a = \left\{\sum_{i=1}^{n} \lambda_i a_i \mid \lambda_i \in \mathbb{C}, a_1 \text{ is an atom}\right\}$$

of all finite linear combinations of atoms. This space is dense in H^1_S since any convergent infinite series can be approximated by finite ones.

Now we will prove the following main theorem which follows the outline for the case of balls presented in [Ste93].

Theorem 184 (Main Theorem)**.**

(a) Suppose $f \in \mathrm{BMO}_S(\mathbb{R}^n)$*. Then the linear functional given by*

$$\ell(g) = \int_{\mathbb{R}^n} f(x)g(x), \quad g \in H^1_S,$$

initially defined on the dense subspace H^1_a*, has a unique bounded extension to* $H^1_{\mathcal{S}}$ *and satisfies*

$$\|\ell\| \leq c\,\|f\|_{BMO_{\mathcal{S}}}.$$

(b) Conversely, every continuous linear function ℓ *on* $H^1_{\mathcal{S}}$ *can be realized as above, with* $f \in BMO_{\mathcal{S}}(\mathbb{R}^n)$*, and with*

$$\|f\|_{BMO_{\mathcal{S}}} \leq c'\,\|\ell\|.$$

Note that this theorem is telling us that the dual space of $H^1_{\mathcal{S}}$ is $BMO_{\mathcal{S}}(\mathbb{R}^n)$, $(H^1_{\mathcal{S}})^* = BMO_{\mathcal{S}}$, just as in the case with cubes.

Proof. To prove that for every $f \in BMO_{\mathcal{S}}(\mathbb{R}^n)$ the linear functional is bounded on H^1 depends on the inequality

$$\left|\int_{\mathbb{R}}^{n} fg\,dx\right| \leq c\,\|f\|_{BMO_{\mathcal{S}}}\,\|g\|_{H^1_{\mathcal{S}}}$$

for $f \in BMO_{\mathcal{S}}$ and $g \in H^1_a \subset H^1_{\mathcal{S}}$. To prove this we will start by assuming that f is bounded. Then we can write

$$\int_{\mathbb{R}^n} fg\,dx = \sum_k \lambda_k \int_{\mathbb{R}^n} f(x)a_k(x)\,dx$$

where $g = \sum_{k=1}^{N} \lambda_k a_k$ is an atomic decomposition for $g \in H^1_a$. Since each a_k is supported in some shape S_k and $\int_{S_k} a_k\,dx = 0$ we can write

$$\int_{\mathbb{R}^n} f(x)a_k(x)\,dx = \int_{S_k} [f(x) - f_{S_k}]a_k(x)\,dx$$

since f_{S_k} is a constant the right and side of that equation is

$$\int_{S_k} [f(x) - f_{S_k}]a_k(x)\,dx = \int_{S_k} f(x)a_k(x)\,dx - \int_{S_k} f_{S_k}a_k(x)\,dx = \int_{S_k} f(x)a_k(x)\,dx.$$

Using the fact that $|a_k(x)| \leq |S_k|^{-1}$, we have that

$$\left|\int f(x)g(x)\,dx\right| \leq \sum_{k=1}^{N} \frac{|\lambda_k|}{|S_k|}\int_{S_k} \left|f(x) - f_{S_k}\right|dx \leq \sum_{k=1}^{N} |\lambda_k|\,\|f\|_{BMO_{\mathcal{S}}}.$$

Thus we have proven the inequality for bounded f. To prove it for general f, let $g \in H^1_a$ again and we replace f by $f^{(k)}$ where

$$f^{(k)} = \begin{cases} -k & \text{if } f(x) \leq -k \\ f(x) & \text{if } -k \leq f(x) \leq k \\ k & \text{if } k \leq f(x). \end{cases}$$

Then since $\left\|f^{(k)}\right\|_{BMO_{\mathcal{S}}} \leq c\,\|f\|_{BMO_{\mathcal{S}}}$, we get by the case just proved that

$$\left\|f^{(k)}\right\|_{BMO_{\mathcal{S}}} \leq c\,\|f\|_{BMO_{\mathcal{S}}}\,\|g\|_{H^1_a}.$$

Finally, since $f^{(k)}$ tends to f almost everywhere as $k \to \infty$, the dominated convergence theorem gives the inequality. It is here that we need the fact that $f \in H^1_a$ and cannot take general $g \in H^1_S$. We have therefore seen that each $f \in \mathrm{BMO}_S$ gives a bounded linear function on the dense subspace H^1_a and thus it extends to all of H^1_S.

Now we prove the converse. The idea is to find a suitable function f that can serve as the representation of the linear functional on each shape S and then we can glue them together to get a global f that works everywhere. We will then show that this function f is indeed in BMO_S. Fix a shape $S \subset \mathbb{R}^n$, and let

$$L^2_S = \left\{ f \in L^1_{loc} \mid \mathrm{Supp}\, f \subset S, \int_S |f(x)|^2 dx < \infty \right\}.$$

Now let

$$L^2_{S,0} = \left\{ f \in L^2_S \mid \int_S f(x) dx = 0 \right\}.$$

Note that for any $g \in L^2_{S,0}$ we have $\|g\|_{H^1_a} \le |S|^{1/2} \|g\|_{L^2_S}$. Indeed, since $g \in L^2_{S,0}$ we know that $\|g\|_{L^2_S} = \|g\|_2 < \infty$. Let $\lambda = \|g\|_2 |S|^{1/2}$. Then observe that

$$\left\| \frac{g}{\lambda} \right\|_2 = \frac{1}{|S|^{1/2}} \|g\|_2 = \frac{1}{|S|^{1/2}}.$$

Additionally, we have $\mathrm{Supp}\, \frac{g}{\lambda} \subset S$ and $\int_S \frac{g}{\lambda} dx = \frac{1}{\lambda} \int_S g dx = 0$. Thus $\frac{g}{\lambda}$ is an atom, and one representation of it is g/λ. This means that $\|g/\lambda\|_{H^1_a} \le 1$. So we get

$$\|g\|_{H^1_a} = \lambda \left\| \frac{g}{\lambda} \right\|_{H^1_a} \le \lambda = |S|^{1/2} \|g\|_2 ,$$

proving the inequality. This inequality tells us that $L^2_{S,0} \subset H^1_a$. Next, if ℓ is a given linear functional on H^1_a, we will assume that it has norm less than or equal to 1, then ℓ extends to a linear functional on $L^2_{S,0}$ with norm at most $|S|^{1/2}$. By the Riesz representation theorem for the Hilbert space $L^2_{S,0}$, there exists an element $F^S \in L^2_{S,0}$ so that

$$\ell(g) = \int_S F^S(x) g(x) dx, \text{ if } g \in L^2_{S,0},$$

with

$$\left(\int_S \left| F^S(x) \right|^2 dx \right)^{1/2} \le |S|^{1/2}.$$

Thus for each shape S, we get a function F^S that represents the linear functions on S. We want to combine these in a way to get a single function f so that, on each shape S, f differs from F^S by a constant. To construct this f observe that if $S_1 \subset S_2$ are shapes, then $F^{S_1} - F^{S_2}$ is constant on S_1. Indeed, both F^{S_1} and F^{S_2} give the same functional on $L^2_{S_1,0}$, so they must differ by a constant on S_1. We can modify F^S, replacing it with $f^S = F^S - c_S$ where c_S is a constant chosen so that f^S has integral 0 on the unit ball centered at the origin. It follows that $f^{S_1} = f^{S_2}$ on S_1 if $S_1 \subset S_2$. Therefore we can unambiguously define f on all of $\mathbb{R}^n$ by taking $f(x) = f^S(x)$ for $x \in S$.

Note that

$$\frac{1}{|S|} \int_S |f(x) - c_S| \, dx \le \frac{1}{|S|} |S|^{1/2} \left(\int_S |f(x) - c_S|^2 \, dx \right)^{1/2} = \left(\frac{1}{|S|} \int_S \left| F^S \right|^2 dx \right)^{1/2} \le 1.$$

Therefore, $f \in BMO_{\mathcal{S}}$ with $\|f\|_{BMO_{\mathcal{S}}} \leq 1$. Also, this gives

$$\ell(g) = \int_{\mathbb{R}^n} f(x)g(x)dx$$

whenever $g \in L^2_{S,0}$ for some S, in particular this representation holds for all $g \in H^1_a$. Thus (b) is proved. $\square$

Bibliography

[AF08] Collin Adams and Robert Franzosa. *Introduction to Topology: Pure and Applied*. Prentice Hall, 2008.

[Bas13] Richard Bass. *Real Analysis for Graduate Students*. 2013.

[Con90] John B. Conway. *A Course in Functional Analysis, Second Edition*. Springer-Verlag, 1990.

[DG20] Galia Dafni and Ryan Gibara. BMO on shapes and sharp constants. In *Advances in harmonic analysis and partial differential equations*, volume 748 of *Contemp. Math.*, pages 1–33. Amer. Math. Soc., [Providence], RI, [2020] ©2020.

[Fef71] Charles Fefferman. Characterizations of bounded mean oscillation. *Bull. Amer. Math. Soc.*, 77:587–588, 1971.

[Fol99] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

[Gra14a] Loukas Grafakos. *Classical Fourier analysis*, volume 249 of *Graduate Texts in Mathematics*. Springer, New York, third edition, 2014.

[Gra14b] Loukas Grafakos. *Modern Fourier Analysis, Third Edition*. Springer-Verlag, 2014.

[Kre78] Erwin Kreyszig. *Introductory Functional Analysis with Applications*. John Wiley & Sons. Inc., 1978.

[Lan87] Serge Lang. *Linear Algebra, third edition*. Springer-Verlag, 1987.

[Roy88] H. L. Royden. *Real Analysis, third edition*. Macmillan Publishing Company, 1988.

[Rud76] Walter Rudin. *Principles of Mathematical Analysis, third edition*. McGraw-Hill, 1976.

[Rud87] Walter Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[Ste93] Elias M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*, volume 43 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1993. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III.

[Tao11] Terence Tao. *An Introduction to Measure Theory*. American Mathematical Society, 2011.

[Tao15a] Terence Tao. *Analysis I*. Springer, 2015.

[Tao15b] Terence Tao. *Analysis II*. Springer, 2015.

[Vel06] Daniel J. Velleman. *How to Prove It: A structured Approach, second edition*. Cambridge University Press, 2006.