JUHA KINNUNEN

Measure and Integral

Department of Mathematics, Aalto University

2019

Contents

All sets can be measured by an outer measure, which is monotone and countably subadditive function. Most important example is the Lebesgue outer measure, which generalizes the concept of volume to all sets. An outer measure has a proper measure theory on measurable sets. A set is Lebesgue measurable if it is almost a Borel set. Existence of a nonmeasurable set for the Lebesgue outer measure is shown by the axiom of choice.

1 Measure theory

1.1 Outer measures

We begin with a general definition of outer measure Let X be a set and consider a mapping on the collection of subsets of X. Recall that if $A_i \subset X$ for $i = 1, 2, \dots$, then

$$\bigcup_{i=1}^{\infty} A_i = \{x \in X : x \in A_i \text{ for some } i\}$$

and

$$\bigcap_{i=1}^{\infty} A_i = \{x \in X : x \in A_i \text{ for every } i\}.$$

Moreover $X \setminus A = \{x \in X : x \notin A\}$.

Definition 1.1. A mapping $\mu^* : \{A : A \subset X\} \to [0, \infty]$ is an outer measure on X, if

(1) $\mu^*(\emptyset) = 0$,

(2) (monotonicity) $\mu^*(A) \leqslant \mu^*(B)$ whenever $A \subset B \subset X$ and

(3) (countable subadditivity) $\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mu^*(A_i)$.

THE MORAL: An outer measure is a countably subadditive set function. Countable subadditivity implies finite subdditivity, since we may add countably many empty sets.

WARNING: It may happen that the equality $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ fails for A and B with $A \cap B = \emptyset$. This means that an outer measure is not necessarily additive on pairwise disjoint sets. Observe that $\leqslant$ holds by subadditivity.

Example 1.2. Let $X = \{1, 2, 3\}$ and define $\mu^*(\emptyset) = 0$, $\mu^*(X) = 2$ and $\mu^*(E) = 1$ for all other $E \subset X$. Then μ^* is an outer measure on X. However, if $A = \{1\}$ and $B = \{2\}$, then

$$\mu^*(A \cup B) = \mu^*(\{1, 2\}) = 1 \neq 2 = \mu^*(A) + \mu^*(B).$$

Figure 1.1: Countable subadditivity.

Examples 1.3:

(1) (The trivial measure) Let $\mu^*(A) = 0$ for every $A \subset X$. Then μ^* is an outer measure. The trivial measure is relatively useless, since all sets have measure zero.

(2) (The discrete measure) Let

$$\mu^*(A) = \begin{cases} 1, & A \neq \emptyset, \\ 0, & A = \emptyset. \end{cases}$$

The discrete outer measure tells whether or not a set is empty.

(3) (The Dirac measure) Let $x_0 \in X$ be a fixed point and let

$$\mu^*(A) = \begin{cases} 1, & x_0 \in A, \\ 0, & x_0 \notin A. \end{cases}$$

This is called the Dirac outer measure at x_0. The Dirac measure tells whether or not a set contains the point x_0.

(4) (The counting measure) Let $\mu^*(A)$ be the (possibly infinite) number of points in A. The counting outer measure tells the number of points of a set.

(5) (The Lebesgue measure) Let $X = \mathbb{R}^n$ and consider the n-dimensional interval

$$I = \{x \in \mathbb{R}^n : a_i \leqslant x_i \leqslant b_i, i = 1, \dots, n\} = [a_1, b_1] \times \dots \times [a_n, b_n]$$

with sides parallel to the coordinate axes. We allow intervals to be degenerate, that is, $b_i = a_i$ for some i. The geometric volume of I is

$$\mathrm{vol}(I) = (b_1 - a_1)(b_2 - a_2)\cdots(b_n - a_n).$$

The Lebesgue outer measure of a set $A \subset \mathbb{R}^n$ is defined as

$$m^*(A) = \inf\left\{\sum_{i=1}^{\infty} \mathrm{vol}(I_i) : A \subset \bigcup_{i=1}^{\infty} I_i\right\},$$

where the infimum is taken over all coverings of A by countably many intervals I_i, $i = 1, 2, \dots$. Observe that this includes coverings with a finite number of intervals, since we may add countably many intervals $I_i = \{x_i\}$ containing only one point with $\mathrm{vol}(I_i) = 0$. The Lebesgue outer measure is nonnegative but may be infinite, so that $0 \leqslant m^*(A) \leqslant \infty$. By the definition of infimum, for every $\varepsilon > 0$, there are intervals I_i, $i = 1, 2, \dots$, such that $A \subset \bigcup_{i=1}^{\infty} I_i$ and

$$m^*(A) \leqslant \sum_{i=1}^{\infty} \mathrm{vol}(I_i) < m^*(A) + \varepsilon.$$

THE MORAL: The Lebesgue outer measure of a set is the infimum of sums of volumes of countably many intervals that cover the set.

We shall discuss more about the Lebesgue outer measure later, but it generalizes the notion of n-dimensional volume to arbitrary subsets of $\mathbb{R}^n$.

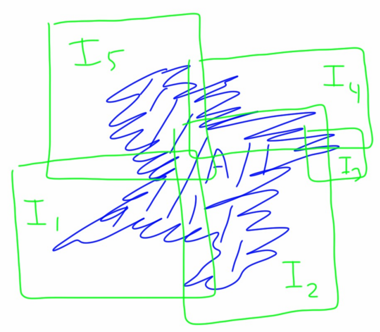

Figure 1.2: Covering by intervals.

Claim: m^* is an outer measure

Reason. $\boxed{(1)}$ Let $\varepsilon > 0$. Since $\emptyset \subset [-\varepsilon^{1/n}/2, \varepsilon^{1/n}/2]^n$, we have

$$0 \leqslant m^*(\emptyset) \leqslant \mathrm{vol}([-\varepsilon^{1/n}/2, \varepsilon^{1/n}/2]^n) = (2\varepsilon^{1/n}/2)^n = \varepsilon.$$

By letting $\varepsilon \to 0$, we conclude $m^*(\emptyset) = 0$. We could also cover $\emptyset$ by the degenerate interval $[x_1, x_1] \times \cdots \times [x_n, x_n]$ for any $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ and conclude the claim from this.

$\boxed{(2)}$ Let $A \subset B$. We may assume that $m^*(B) < \infty$, for otherwise $m^*(A) \leqslant m^*(B) = \infty$ and there is nothing to prove. For every $\varepsilon > 0$ there are intervals I_i, $i = 1, 2, \ldots$, such that $B \subset \bigcup_{i=1}^{\infty} I_i$ and

$$\sum_{i=1}^{\infty} \mathrm{vol}(I_i) < m^*(B) + \varepsilon.$$

Since $A \subset B \subset \bigcup_{i=1}^{\infty} I_i$, we have

$$m^*(A) \leqslant \sum_{i=1}^{\infty} \mathrm{vol}(I_i) < m^*(B) + \varepsilon.$$

By letting $\varepsilon \to 0$, we conclude $m^*(A) \leqslant m^*(B)$.

$\boxed{(3)}$ We may assume that $m^*(A_i) < \infty$ for every $i = 1, 2, \ldots$, for otherwise there is nothing to prove. Let $\varepsilon > 0$. For every $i = 1, 2, \ldots$ there are intervals $I_{j,i}$, $j = 1, 2, \ldots$, such that $A_i \subset \bigcup_{j=1}^{\infty} I_{j,i}$ and

$$\sum_{j=1}^{\infty} \mathrm{vol}(I_{j,i}) < m^*(A_i) + \frac{\varepsilon}{2^i}.$$

Then $\bigcup_{i=1}^{\infty} A_i \subset \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} I_{j,i} = \bigcup_{i,j=1}^{\infty} I_{j,i}$ and

$$\begin{aligned} m^*\left(\bigcup_{i=1}^{\infty} A_i\right) &\leqslant \sum_{i,j=1}^{\infty} \mathrm{vol}(I_{j,i}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mathrm{vol}(I_{j,i}) \\ &\leqslant \sum_{i=1}^{\infty} \left(m^*(A_i) + \frac{\varepsilon}{2^i}\right) = \sum_{i=1}^{\infty} m^*(A_i) + \varepsilon. \end{aligned}$$

The claim follows by letting $\varepsilon \to 0$. ∎

(6) (The Hausdorff measure) Let $X = \mathbb{R}^n$, $0 < s < \infty$ and $0 < \delta \leqslant \infty$. Define

$$\mathscr{H}^s_\delta(A) = \inf\left\{\sum_{i=1}^{\infty} (\mathrm{diam}(B_i))^s : A \subset \bigcup_{i=1}^{\infty} B_i,\ \mathrm{diam}(B_i) \leqslant \delta\right\}$$

and

$$\mathscr{H}^s(A) = \lim_{\delta \to 0} \mathscr{H}^s_\delta(A) = \sup_{\delta > 0} \mathscr{H}^s_\delta(A).$$

We call $\mathscr{H}^s$ the s-dimensional Hausdorff outer measure on $\mathbb{R}^n$. This generalizes the notion of s-dimensional measure to arbitrary subsets of $\mathbb{R}^n$.

See Remark 1.23 for the definition of the diameter of a set. We refer to [2, Section 3.8], [4, Chapter 2], [8, Chapter 19], [11, Chapter 7] and [16, Chapter 7] for more.

Observe that, for every $\delta > 0$, an arbitrary set $A \subset \mathbb{R}^n$ can be covered by $B(x, \frac{\delta}{2})$ with $x \in A$, that is,

$$A \subset \bigcup \left\{ B\left(x, \frac{\delta}{2}\right) : x \in A \right\},$$

where $B(x,r) = \{y \in \mathbb{R}^n : |y-x| < r\}$ denotes an open ball of radius $r > 0$ and center x. By Lindelöf's theorem every open covering in $\mathbb{R}^n$ has a countable subcovering. This implies that there exist countably many balls $B_i = B(x_i, \frac{\delta}{2})$, $i = 1,2,\dots$, such that $A \subset \bigcup_{i=1}^\infty B_i$. Moreover, we have $\operatorname{diam}(B_i) \leqslant \delta$ for every $i = 1,2,\dots$. This shows that the coverings in the definition of the Hausdorff outer measure exist.

(7) Let $\mathscr{F}$ be a collection of subsets of X such that $\emptyset \in \mathscr{F}$ and there exist $A_i \in \mathscr{F}$, $i = 1,2,\dots$, such that $X = \bigcup_{i=1}^\infty A_i$. Let $\rho : \mathscr{F} \to [0,\infty]$ be any function for which $\rho(\emptyset) = 0$. Then $\mu^* : \mathscr{P}(X) \to [0,\infty]$,

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^\infty \rho(A_i) : A_i \in \mathscr{F}, A \subset \bigcup_{i=1}^\infty A_i \right\}$$

is an outer measure on X. Moreover, if ρ is monotone and countably subadditive on $\mathscr{F}$, then $\mu^* = \rho$ on $\mathscr{F}$ (exercise). This gives a very general method to construct outer measures, see [2, Section 2.8].

(8) (Carathéodory's construction) Let $X = \mathbb{R}^n$, $\mathscr{F}$ be a collection of subsets of X and $\rho : \mathscr{F} \to [0,\infty]$ be any function. We make the following two assumptions.

- For every $\delta > 0$ there are $A_i \in \mathscr{F}$, $i = 1,2,\dots$, such that $X = \bigcup_{i=1}^\infty A_i$ and $\operatorname{diam}(A_i) \leqslant \delta$.
- For every $\delta > 0$ there is $A \in \mathscr{F}$ such that $\rho(A) \leqslant \delta$ and $\operatorname{diam}(A) \leqslant \delta$.

For $0 < \delta \leqslant \infty$ and $A \subset X$, we define

$$\mu^*_\delta(A) = \inf \left\{ \sum_{i=1}^\infty \rho(A_i) : A_i \in \mathscr{F}, A \subset \bigcup_{i=1}^\infty A_i,\ \operatorname{diam}(A_i) \leqslant \delta \right\}.$$

The first assumption guarantees that we can cover any set A with sets in $\mathscr{F}$ and the second assumption implies $\mu^*_\delta(\emptyset) = 0$. It can be shown that μ^*_δ is an outer measure (exercise), but it is usually not additive on disjoint sets (see Theorem 1.10) and not a Borel outer measure (see Defintion 1.33). Clearly,

$$\mu^*_{\delta'}(A) \leqslant \mu^*_\delta(A) \quad \text{for} \quad 0 < \delta < \delta' \leqslant \infty.$$

Thus we may define

$$\mu^*(A) = \lim_{\delta \to 0} \mu^*_\delta(A) = \sup_{\delta > 0} \mu^*_\delta(A).$$

The outer measure μ^* has much better properties than μ^*_δ. For example, it is always a Borel outer measure (see Theorem 1.48 and Remarks 1.49). Moreover, if the members of $\mathscr{F}$ are Borel sets, then μ^* is Borel regular (see Definition 1.33). This gives a very general method to construct Borel outer measures, see [2, Section 3.3].

THE MORAL: The examples above show that it is easy to construct outer measures. However, we have to restrict ourselves to a class of measurable sets in order to obtain a useful theory.

1.2 Measurable sets

We discuss so-called Carathéodory criterion for measurability for a general outer measure. The definition is perhaps not very intuitive, but it will be useful in the arguments. Later we give a more geometric and intuitive characterizations of measurable sets for the Lebesgue and other outer measures.

Definition 1.4. A set $A \subset X$ is μ^*-measurable, if

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

for every $E \subset X$.

THE MORAL: A measurable set divides an arbitrary set in two parts in an additive way. In practice it is difficult to show directly from the definition that a set is measurable.

Remarks 1.5:

(1) Since $E = (E \cap A) \cup (E \setminus A)$, by subadditivity

$$\mu^*(E) \leqslant \mu^*(E \cap A) + \mu^*(E \setminus A).$$

This means that $\leqslant$ holds always in Definition 1.4.

(2) If A is μ^*-measurable and $A \subset B$, where B is an arbitrary subset of X, then

$$\mu^*(B) = \mu^*(\underbrace{B \cap A}_{=A}) + \mu^*(B \setminus A) = \mu^*(A) + \mu^*(B \setminus A).$$

This shows that an outer measure behaves additively on measurable subsets.

(3) If $\mu^*(A) = 0$, then A is μ^*-measurable. In other words, all sets of measure zero are always measurable.

Reason. Since $E \cap A \subset A$ and $E \setminus A \subset E$, we have

$$\mu^*(E \cap A) + \mu^*(E \setminus A) \leqslant \underbrace{\mu^*(A)}_{=0} + \mu^*(E) = \mu^*(E)$$

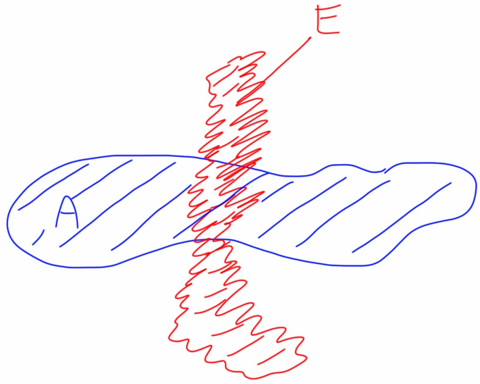

Figure 1.3: A measurable set.

for every $E \subset X$. On the other hand, by (1) we always have inequality in the other direction, so that equality holds. ∎

(4) $\emptyset$ and X are μ^*-measurable. In other words, the empty set and the entire space are always measurable.

Reason.

$$\mu^*(E) = \mu^*(\underbrace{E \cap \emptyset}_{=\emptyset}) + \mu^*(\underbrace{E \setminus \emptyset}_{=E})$$

and

$$\mu^*(E) = \mu^*(\underbrace{E \cap X}_{=E}) + \mu^*(\underbrace{E \setminus X}_{=\emptyset}),$$

for every $E \subset X$. This shows that $\emptyset$ and X are μ^*-measurable. ∎

Another way is to apply Lemma 1.9, which asserts that A is μ^*-measurable if and only if $X \setminus A$ is μ^*-measurable. Hence $X = X \setminus \emptyset$ is μ^*-measurable, since $\emptyset$ is μ^*-measurable as a set of measure zero.

(5) The only measurable sets for the discrete measure are $\emptyset$ and X (exercise). In this case there are extremely few measurable sets.

(6) All sets are measurable for the Dirac measure (exercise). In this case there are extremely many measurable sets.

Example 1.6. (Continuation of Example 1.2) Let $X = \{1,2,3\}$ and, define an outer measure μ^* such that $\mu^*(\emptyset) = 0$, $\mu^*(X) = 2$ and $\mu^*(E) = 1$ for all other $E \subset X$. If $a, b \in X$ are different points, $A = \{a\}$ and $E = \{a, b\}$, then

$$\mu^*(E) = \mu^*(\{a,b\}) = 1 < 2 = \mu^*(\{a\}) + \mu^*(\{b\}) = \mu^*(E \cap A) + \mu^*(E \setminus A).$$

This means that A is not μ^*-measurable. In the same way we can see that all sets consisting of two points are not μ^*-measurable. Thus only μ^*-measurable sets are $\emptyset$ and X.

Next we discuss the structural properties of measurable sets.

Definition 1.7. A collection $\mathscr{A}$ of subsets of X is a σ-algebra, if

(1) $\emptyset \in \mathscr{A}$,
(2) $A \in \mathscr{A}$ implies $A^{\complement} = X \setminus A \in \mathscr{A}$ and
(3) $A_i \in \mathscr{A}$ for every $i = 1, 2, \ldots$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$.

THE MORAL: The letter σ stands for countable and condition (3) allows for countable unions.

Remark 1.8. Every σ-algebra $\mathscr{A}$ has the following properties.

(1) Since $X \setminus \emptyset = X$, by (1) and (2) in Defintion 1.7 we have $X \in \mathscr{A}$.
(2) If $A_1, \ldots, A_k \in \mathscr{A}$, then $\bigcup_{i=1}^{k} A_i \in \mathscr{A}$. This follows from (3) in Definition 1.7 by taking $A_i = \emptyset$ for $i = k+1, k+2, \ldots$.
(3) If $A_i \in \mathscr{A}$ for every $i = 1, 2, \ldots$, then $\bigcap_{i=1}^{\infty} A_i \in \mathscr{A}$. To see this, observe that by de Morgan's law and (3) in Definition 1.7

$$X \setminus \bigcap_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} (X \setminus A_i) \in \mathscr{A}.$$

By taking $A_i = \emptyset$ for $i = k+1, k+2, \ldots$, we also have $\bigcap_{i=1}^{k} A_i \in \mathscr{A}$.
(4) If $A, B \in \mathscr{A}$, by (2) in Definition 1.7 and remark (2) above, we have $A \setminus B = A \cap (X \setminus B) \in \mathscr{A}$.

Lemma 1.9. The collection $\mathscr{M}$ of μ^*-measurable sets is a σ-algebra.

THE MORAL: The collection of measurable sets is closed under countably many set theoretic operations of taking complements, unions and intersections.

Proof. $\boxed{(1)}$ $\mu^*(\emptyset) = 0$ implies that $\emptyset \in \mathscr{M}$, see Remark 1.5.

$\boxed{(2)}$ $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) = \mu^*(E \setminus (X \setminus A)) + \mu^*(E \cap (X \setminus A))$ for every $E \subset X$. This implies that $A^{\complement} \in \mathscr{M}$.

$\boxed{(3)}$ **Step 1:** First we show that $A_1, A_2 \in \mathcal{M}$ implies $A_1 \cup A_2 \in \mathcal{M}$.

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \setminus A_1) + \mu^*(E \cap A_1) \quad (A_1 \in \mathcal{M}, E \text{ test set}) \\
&= \mu^*((E \setminus A_1) \cap A_2) + \mu^*((E \setminus A_1) \setminus A_2) + \mu^*(E \cap A_1) \\
&\qquad (A_2 \in \mathcal{M}, E \setminus A_1 \text{ as a test set}) \\
&= \mu^*((E \setminus A_1) \cap A_2) + \mu^*(E \setminus (A_1 \cup A_2)) + \mu^*(E \cap A_1) \\
&\qquad ((E \setminus A_1) \setminus A_2 = E \setminus (A_1 \cup A_2)) \\
&\geqslant \mu^*(((E \setminus A_1) \cap A_2) \cup (E \cap A_1)) + \mu^*(E \setminus (A_1 \cup A_2)) \quad \text{(subadditivity)} \\
&= \mu^*(E \cap (A_1 \cup A_2)) + \mu^*(E \setminus (A_1 \cup A_2)) \\
&\qquad ((E \setminus A_1) \cap A_2) \cup (E \cap A_1) = E \cap (A_1 \cup A_2))
\end{aligned}$$

for every $E \subset X$. By iteration, the same result holds for finitely many sets: If $A_i \in \mathcal{M}, i = 1, 2, \dots, k$, then $\bigcup_{i=1}^k A_i \in \mathcal{M}$. By de Morgan's law, we also have $\bigcap_{i=1}^k A_i \in \mathcal{M}$.

Step 2: We construct pairwise disjoint sets C_i such that $C_i \subset A_i$ and $\bigcup_{i=1}^\infty A_i = \bigcup_{i=1}^\infty C_i$. Let $B_k = \bigcup_{i=1}^k A_i$, $k = 1, 2, \dots$ Then $B_k \subset B_{k+1}$ and

$$\bigcup_{i=1}^\infty A_i = B_1 \cup \left(\bigcup_{k=1}^\infty (B_{k+1} \setminus B_k) \right).$$

Let

$$C_1 = B_1 \quad \text{and} \quad C_{i+1} = B_{i+1} \setminus B_i, \quad i = 1, 2, \dots$$

Then $C_i \cap C_j = \emptyset$ whenever $i \neq j$ and the sets C_i, $i = 1, 2, \dots$, have the required properties. The sets $C_i \in \mathcal{M}$, since they are finite unions and intersections μ^*-measurable sets, see Step 1.

Step 3: By the argument in Step 2 we may assume that the sets $A_i \in \mathcal{M}$, $i = 1, 2, \dots$, are pairwise disjoint, that is, $A_i \cap A_j = \emptyset$ for $i \neq j$. Let $B_k = \bigcup_{i=1}^k A_i$, $k = 1, 2, \dots$ We show by induction that

$$\mu^*(E \cap B_k) = \sum_{i=1}^k \mu^*(E \cap A_i), \quad k = 1, 2, \dots,$$

for every $E \subset X$.

Note: By choosing $E = X$, this implies finite additivity on disjoint measurable sets. Observe, that $\leqslant$ holds by subadditivity.

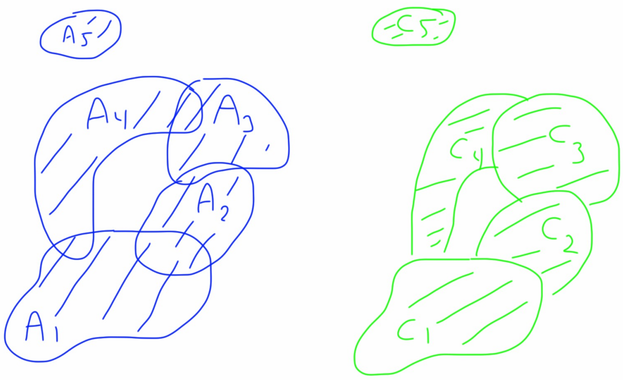

Figure 1.4: Covering by disjoint sets.

The case $k = 1$ is clear. Assume that the claim holds with index k. Then

$$\begin{aligned}
\mu^*(E \cap B_{k+1}) &= \mu^*((E \cap B_{k+1}) \cap B_k) + \mu^*((E \cap B_{k+1}) \setminus B_k) \\
&\qquad (B_k \in \mathscr{M}, E \cap B_{k+1} \text{ as a test set}) \\
&= \mu^*(E \cap B_k) + \mu^*(E \cap A_{k+1}) \\
&\qquad (B_k \subset B_{k+1},\ A_i \text{ are pairwise disjoint implies } B_{k+1} \setminus B_k = A_{k+1}) \\
&= \sum_{i=1}^{k} \mu^*(E \cap A_i) + \mu^*(E \cap A_{k+1}) \quad \text{(the induction assumption)} \\
&= \sum_{i=1}^{k+1} \mu^*(E \cap A_i).
\end{aligned}$$

Step 4: By Step 3 and monotonicity with $B_k \subset \bigcup_{i=1}^{\infty} A_i$, we have

$$\sum_{i=1}^{k} \mu^*(E \cap A_i) = \mu^*(E \cap B_k) \leqslant \mu^* \left(E \cap \bigcup_{i=1}^{\infty} A_i \right).$$

This implies

$$\sum_{i=1}^{\infty} \mu^*(E \cap A_i) = \lim_{k \to \infty} \sum_{i=1}^{k} \mu^*(E \cap A_i) \leqslant \mu^* \left(E \cap \bigcup_{i=1}^{\infty} A_i \right).$$

On the other hand, by subadditivity

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

This shows that

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

whenever $A_i\in\mathcal{M}$, $i=1,2,\dots$, are disjoint.

Note: By choosing $E=X$, this implies countable additivity on pairwise disjoint measurable sets.

Step 5: Let $E\subset X$, $A=\bigcup_{i=1}^{\infty}A_i$ with pairwise disjoint $A_i\in\mathcal{M}$, $i=1,2,\dots$. Then

$$\begin{aligned}\mu^*(E)&=\mu^*(E\cap B_k)+\mu^*(E\setminus B_k)\quad(\text{Step 1 implies } B_k\in\mathcal{M})\\&=\sum_{i=1}^{k}\mu^*(E\cap A_i)+\mu^*(E\setminus B_k)\quad(\text{Step 3})\\&\geqslant\sum_{i=1}^{k}\mu^*(E\cap A_i)+\mu^*(E\setminus A),\quad k=1,2,\dots.\quad(B_k\subset A)\end{aligned}$$

This implies

$$\begin{aligned}\mu^*(E)&\geqslant\lim_{k\to\infty}\sum_{i=1}^{k}\mu^*(E\cap A_i)+\mu^*(E\setminus A)\\&=\sum_{i=1}^{\infty}\mu^*(E\cap A_i)+\mu^*(E\setminus A)\\&=\mu^*(E\cap A)+\mu^*(E\setminus A).\quad(\text{Step 4})\end{aligned}$$

Note that Step 4 is not really needed here. We could have used countable subadditivity instead as an inequality is enough here. However, we need equality in Step 4 in the proof of Theorem 1.10. □

1.3 Measures

From the proof of Lemma 1.9 we see that an outer measure is countably additive on pairwise disjoint measurable sets. This is a very useful property. Example 1.2 shows this does not necessarily hold for sets that are not measurable. The overall idea is that an outer measure produces a proper measure theory when restricted to measurable sets.

Theorem 1.10. (Countable additivity) Assume that $A_i\subset X$, $i=1,2,\dots$, are pairwise disjoint ($A_i\cap A_j=\emptyset$ for $i\neq j$) and μ^*-measurable set. Then

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

THE MORAL: An outer measure is countably additive on pairwise disjoint measurable sets. The measure theory is compatible under partitions a given measurable set into countably many pairwise disjoint measurable sets.

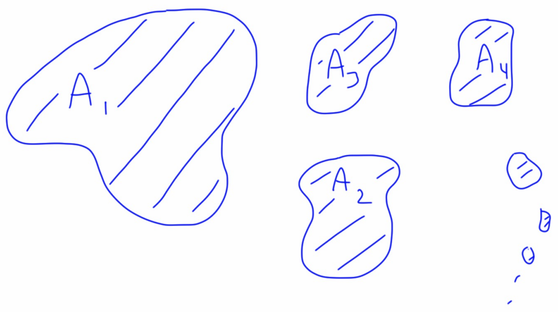

Figure 1.5: Disjoint sets.

Proof. By Step 4 of the the proof of Lemma 1.9

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

whenever $A_i \in \mathscr{M}$, $i = 1,2,\dots$, are pairwise disjoint μ^*-measurable and $E \subset X$. The claim follows by choosing $E = X$. □

Definition 1.11. Assume that $\mathscr{M}$ is σ-algebra in X. A mapping $\mu\colon \mathscr{M} \to [0,\infty]$ is a measure on a measure space $(X,\mathscr{M},\mu)$, if

(1) $\mu(\emptyset) = 0$ and

(2) $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $A_i \in \mathscr{M}$, $i = 1,2,\dots$, are pairwise disjoint.

THE MORAL: A measure is a countably additive set function on pairwise disjoint sets in the σ-algebra. An outer measure is defined on all subsets, but a measure is defined only on sets in the σ-algebra.

Remarks 1.12:

(1) A measure μ is monotone on $\mathscr{M}$, since

$$\mu(B) = \mu(A) + \mu(B \setminus A) \geqslant \mu(A)$$

for every $A,B \in \mathscr{M}$ with $A \subset B$. In the same way we can see that μ is countably subadditive on $\mathscr{M}$.

(2) It is possible to develop a theory also for signed or even complex valued measures. Assume that $\mathcal{M}$ is σ-algebra in X. A mapping $\mu\colon \mathcal{M} \to [-\infty,\infty]$ is a signed measure on a measure space $(X,\mathcal{M},\mu)$, if $\mu(\emptyset)=0$ and whenever $A_i \in \mathcal{M}$ are disjoint, then $\sum_{i=1}^{\infty}\mu(A_i)$ exists as an extended real number, tha is the sum converges, and

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

Every subset of a set of outer measure zero is measurable as a set of outer measure zero. In contrast, there is a delicate issue related to sets of measure zero for a measure defined on a σ-algebra.

Definition 1.13. A measure μ on a measure space $(X,\mathcal{M},\mu)$ is said to be complete, if $B \in \mathcal{M}$, $\mu(B)=0$ and $A \subset B$ implies $A \in \mathcal{M}$.

THE MORAL: A measure is complete, if every subset of a set of measure zero is measurable. This will be useful when we discuss properties that hold outside sets of measure zero, see Remark 2.30.

Example 1.14. It is possible that $A \subset B \in \mathcal{M}$ and $\mu(B)=0$, but $A \notin \mathcal{M}$.

Reason. Let $X=\{1,2,3\}$ and $\mathcal{M}=\{\emptyset,\{1\},\{2,3\},X\}$. Then σ is a σ-algebra. Define a measure on $\mathcal{M}$ by $\mu(\emptyset)=0$, $\mu(\{1\})=1$, $\mu(\{2,3\})=0$ and $\mu(X)=1$. In this case $\{2,3\}\in\mathcal{M}$ and $\mu^*(\{2,3\})=0$, but $\{2\}\notin\mathcal{M}$. ■

Remark 1.15. For example, the measure space $(\mathbb{R}^n,\mathcal{B},\mu)$, where $\mathcal{B}$ denotes the Borel sets and μ the Lebesgue measure is not complete, see Definition 1.31 and discussion in Section 2.3. Every measure space can be completed in a natural way by adding all sets of measure zero to the σ-algebra (exercise). See also [2, Theorem 2.26], [11, Exercise 2, p. 312] and [12, Exercise 1.4.6, p. 78].

The following finiteness condition is useful for us later.

Definition 1.16. A measure μ on a measure space $(X,\mathcal{M},\mu)$ is σ-finite if $X = \bigcup_{i=1}^{\infty} A_i$, where $A_i \in \mathcal{M}$ and $\mu(A_i)<\infty$ for every $i=1,2,\dots$.

THE MORAL: If a measure is σ-finite, then the entire space can be covered by measurable sets with finite measure. The corresponding notion can be defined for outer measures as well.

Lemma 1.17. The Lebesgue outer measure m^* is σ-finite.

Proof. Clearly $\mathbb{R}^n = \bigcup_{i=1}^{\infty} B(0,i)$, where $B(0,i) = \{x \in R^n : |x| < i\}$ is a ball with center at the origin and radius i. The Lebesgue outer measure of the ball $B(0,i)$ is finite, since

$$m^*(B(0,i)) \leqslant \text{vol}([-i,i]^n) = (2i)^n < \infty.$$

We shall show later that all open sets are Lebesgue measurable, see Lemma 1.50. □

Remark 1.18. Every outer measure restricted to measurable sets induces a complete measure. On the other hand, every measure on a measure space $(X, \mathscr{M}, \mu)$ induces an outer measure

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \mu(A) : E \subset \bigcup_{i=1}^{\infty} A_i, A_i \in \mathscr{M} \text{ for every } i = 1,2,\dots\right\}.$$

Assume that the measure space $(X, \mathscr{M}, \mu)$ is σ-finite, see Definition 1.16. Then every set in $\mathscr{M}$ is μ^*-measurable and $\mu = \mu^*$ on $\mathscr{M}$. This means that μ^* is an extension of μ. If the measure space $(X, \mathscr{M}, \mu)$ is complete, then the class of μ^*-measurable sets is precisely $\mathscr{M}$ (exercise). See also [2, p. 99–116], [8, Lemma 9.6], [11, p. 270–273] and [12, p. 153–156].

THE MORAL: If the measure is σ-finite and complete, then there will be no new measurable sets when we switch to the induced outer measure. In this sense the difference between outer measures and measures is small.

Next we give some examples of measures.

Examples 1.19:

(1) $(X, \mathscr{M}, \mu)$, where X is a set, μ is an outer measure on X and $\mathscr{M}$ is the σ-algebra of μ-measurable sets.

(2) $(\mathbb{R}^n, \mathscr{M}, m^*)$, where m^* is the Lebesgue outer measure and $\mathscr{M}$ is the σ-algebra of Lebesgue measurable sets.

(3) A measure space $(X, \mathscr{M}, \mu)$ with $\mu(X) = 1$ is called a probability space, μ a probability measure and sets belonging to $\mathscr{M}$ events.

The next theorem shows that an outer measure has useful monotone convergence properties on measurable sets.

Theorem 1.20. Assume that μ^* is an outer measure on X and that sets $A_i \subset X$, $i = 1,2,\dots$, are μ^*-measurable.

(1) (Upwards monotone convergence) If $A_1 \subset A_2 \subset \cdots$, then

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

(2) (Downwards monotone convergence) If $A_1 \supset A_2 \supset \cdots$, and $\mu^*(A_{i_0}) < \infty$ for some i_0, then

$$\lim_{i\to\infty} \mu^*(A_i) = \mu^*\left(\bigcap_{i=1}^{\infty} A_i\right).$$

THE MORAL: The measure theory is compatible under taking limits, if we approximate a given measurable set with an increasing sequence of measurable sets from inside or a decreasing sequence of measurable sets from outside.

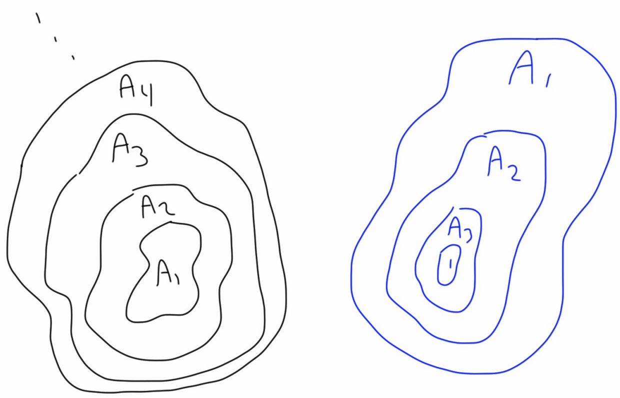

Figure 1.6: Monotone sequences of sets.

Remarks 1.21:

(1) The results do not hold, in general, without the measurability assumptions.

Reason. Let $X = \mathbb{N}$. Define an outer measure on $\mathbb{N}$ by

$$\mu^*(A) = \begin{cases} 0, & A = \emptyset, \\ 1, & A \text{ finite}, \\ 2, & A \text{ infinite}. \end{cases}$$

Let $A_i = \{1,2,\dots,i\}$, $i = 1,2,\dots$ Then

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) = 2 \neq 1 = \lim_{i\to\infty} \mu^*(A_i).$$

■

(2) The assumption $\mu^*(A_{i_0}) < \infty$ is essential in (2).

Reason. Let $X = \mathbb{R}$, m^* be the Lebesgue outer measure and $A_i = [i, \infty)$, $i = 1, 2, \ldots$. Then $\bigcap_{i=1}^{\infty} A_i = \emptyset$ and $m^*(A_i) = \infty$ for every $i = 1, 2, \ldots$ In this case

$$\lim_{i\to\infty} m^*(A_i) = \infty, \text{ but } m^*\left(\bigcap_{i=1}^{\infty} A_i\right) = m^*(\emptyset) = 0.$$

■

(3) The following observation will be used several times in the proof of Theorem 1.20. Assume that A is μ^*-measurable and let $B \subset \mathbb{R}^n$ be any set with $A \subset B$ and $\mu^*(A) < \infty$. By Definition 1.4, we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) = \mu^*(A) + \mu^*(B \setminus A)$$

and thus $\mu^*(B \setminus A) = \mu^*(B) - \mu(A)$. If both A and B are μ^*-measurable, we can conclude the same result from additivity on disjoint measurable sets as

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

Proof. $\boxed{(1)}$ We may assume that $\mu^*(A_i) < \infty$ for every i, otherwise the claim follows from monotonicity.

$$\begin{aligned}
\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mu^*\left(A_1 \cup \bigcup_{i=1}^{\infty} (A_{i+1} \setminus A_i)\right) \\
&= \mu^*(A_1) + \sum_{i=1}^{\infty} \mu^*(A_{i+1} \setminus A_i) \quad \text{(disjointness and measurability)} \\
&= \mu^*(A_1) + \sum_{i=1}^{\infty} \Big(\mu^*(A_{i+1}) - \mu^*(A_i)\Big) \\
&\qquad (\mu^*(A_{i+1}) = \mu^*(A_{i+1} \cap A_i) + \mu^*(A_{i+1} \setminus A_i) \text{ and } A_{i+1} \cap A_i = A_i) \\
&= \lim_{k\to\infty} \left(\mu^*(A_1) + \sum_{i=1}^{k} \Big(\mu^*(A_{i+1}) - \mu^*(A_i)\Big)\right) \\
&= \lim_{k\to\infty} \mu^*(A_{k+1}) = \lim_{i\to\infty} \mu^*(A_i).
\end{aligned}$$

$\boxed{(2)}$ By replacing sets A_i by $A_i \cap A_{i_0}$, we may assume that $\mu^*(A_1) < \infty$. $A_{i+1} \subset A_i$ implies $A_1 \setminus A_i \subset A_1 \setminus A_{i+1}$ for every $i = 1, 2, \ldots$. By (1) we have

$$\begin{aligned}
\mu^*\left(\bigcup_{i=1}^{\infty} (A_1 \setminus A_i)\right) &= \lim_{i\to\infty} \mu^*(A_1 \setminus A_i) \\
&= \lim_{i\to\infty} \big(\mu^*(A_1) - \mu^*(A_i)\big) \\
&\qquad (\mu^*(A_1) = \mu^*(A_1 \cap A_i) + \mu^*(A_1 \setminus A_i)) \\
&= \mu^*(A_1) - \lim_{i\to\infty} \mu^*(A_i).
\end{aligned}$$

On the other hand, as above, we have

$$\mu^*\left(\bigcup_{i=1}^{\infty}(A_1 \setminus A_i)\right) = \mu^*\left(A_1 \setminus \bigcap_{i=1}^{\infty} A_i\right) = \mu^*(A_1) - \mu^*\left(\bigcap_{i=1}^{\infty} A_i\right).$$

This implies

$$\mu^*(A_1) - \mu^*\left(\bigcap_{i=1}^{\infty} A_i\right) = \mu^*(A_1) - \lim_{i\to\infty} \mu^*(A_i).$$

Since $\mu^*(A_1) < \infty$, we may conclude

$$\mu^*\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu^*(A_i).$$

□

1.4 The distance function

The distance function will be a useful tool in the sequel.

Definition 1.22. Let $A \subset \mathbb{R}^n$ with $A \neq \emptyset$. The distance from a point $x \in \mathbb{R}^n$ to A is

$$\operatorname{dist}(x, A) = \inf\{|x - y| : y \in A\}.$$

Remarks 1.23:

(1) The distance between the nonempty sets A, $B \subset \mathbb{R}^n$ is

$$\operatorname{dist}(A, B) = \inf\{|x - y| : x \in A \text{ and } y \in B\}.$$

(2) The diameter of the nonempty set $A \subset \mathbb{R}^n$ is

$$\operatorname{diam}(A) = \sup\{|x - y| : x, y \in A\}.$$

Lemma 1.24. Let $A \subset \mathbb{R}^n$ with $A \neq \emptyset$. For every $x \in \mathbb{R}^n$, there exist a point $x_0 \in \overline{A}$ such that

$$\operatorname{dist}(x, A) = |x_0 - x|.$$

T H E M O R A L : There is a closest point in the closure of the set. If A is closed, then the closest point belongs to A. In general, the closest point is not unique.

Proof. Let $x \in \mathbb{R}^n$. There exists a sequence $y_i \in A$, $i = 1, 2,$ such that

$$\lim_{i\to\infty} |x - y_i| = \operatorname{dist}(x, A).$$

The sequence (y_i) is bounded and by Bolzano-Weierstrass theorem it has a converging subsequence (y_{j_k}) such that $y_{j_k} \to x_0$ as $k \to \infty$ for some $x_0 \in \mathbb{R}^n$. Since $\overline{A}$ is a closed set and $y_{j_k} \in A$ for every k, we have $x_0 \in \overline{A}$. Since $y \mapsto |x - y|$ is a continuous function, we conclude

$$|x - x_0| = \lim_{k\to\infty} |x - y_{j_k}| = \operatorname{dist}(x, A).$$

□

Lemma 1.25. Let $A \subset \mathbb{R}^n$ with $A \neq \emptyset$. Then $|\operatorname{dist}(x,A) - \operatorname{dist}(y,A)| \leqslant |x-y|$ for every $x,y \in \mathbb{R}^n$.

T H E M O R A L : The distance function is a Lipschitz continuous function with the Lipschitz constant one.

Proof. Let $x,y \in \mathbb{R}^n$. By the triangle inequality $|x-z| \leqslant |x-y| + |y-z|$ for every $z \in A$. For every $\varepsilon > 0$ there exists $z' \in A$ such that $|y-z'| \leqslant \operatorname{dist}(y,A) + \varepsilon$. Thus

$$\operatorname{dist}(x,A) \leqslant |x-z'| \leqslant |x-y| + \operatorname{dist}(y,A) + \varepsilon,$$

which implies

$$\operatorname{dist}(x,A) - \operatorname{dist}(y,A) \leqslant |x-y| + \varepsilon.$$

By switching the roles of x and y, we obtain

$$|\operatorname{dist}(x,A) - \operatorname{dist}(y,A)| \leqslant |x-y| + \varepsilon.$$

This holds for every $\varepsilon > 0$, so that

$$|\operatorname{dist}(x,A) - \operatorname{dist}(y,A)| \leqslant |x-y|. \qquad \square$$

Lemma 1.26. Let $A \subset \mathbb{R}^n$ be an open set with $\partial A \neq \emptyset$. Define

$$A_i = \left\{ x \in A : \operatorname{dist}(x, \partial A) > \frac{1}{i} \right\}, \quad i = 1,2,\dots$$

Then the sets A_i are open, $A_i \subset A_{i+1}$, $i = 1,2,\dots$, and that $A = \bigcup_{i=1}^{\infty} A_i$.

T H E M O R A L : Any open set can be exhausted by an increasing sequence of distance sets.

Proof. Recall that a function is continuous if and only if the preimage of every open set is open. Thus

$$\left\{ x \in A : \operatorname{dist}(x, \partial A) > \frac{1}{i} \right\} = f^{-1}\left(\left(\frac{1}{i}, \infty \right) \right)$$

is an open set. It is immediate that $A_i \subset A_{i+1}$, $i = 1,2,\dots$.

Since $A_i \subset A$ for every $i = 1,2,\dots$, we have $\bigcup_{i=1}^{\infty} A_i \subset A$. On the other hand, since A is open, for every $x \in A$ there exists $\varepsilon > 0$ such that $B(x,\varepsilon) \subset A$. This implies $\operatorname{dist}(x, \partial A) \geqslant \varepsilon$. Thus we may choose i large enough so that $x \in A_i$. This shows that $A \subset \bigcup_{i=1}^{\infty} A_i$. $\square$

Lemma 1.27. If $A \subset \mathbb{R}^n$ is an open with $\partial A \neq \emptyset$, then $\operatorname{dist}(K, \partial A) > 0$ for every compact subset K of A.

Proof. Since $x \mapsto \operatorname{dist}(x, \partial A)$ is a continuous function, it attains its minimum on any compact set. Thus there exists $z \in K$ such that $\operatorname{dist}(z, \partial A) = \operatorname{dist}(K, \partial A)$. Since A is open and $z \in A$, there exists $\varepsilon > 0$ such that $B(z, \varepsilon) \subset A$. This implies

$$\operatorname{dist}(K, \partial A) = \operatorname{dist}(z, \partial A) \geqslant \varepsilon > 0. \qquad \square$$

WARNING: The corresponding claim does not hold if $K \subset A$ only assumed to be closed. For example, $A = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ is open, $K = \{(x, y) \in \mathbb{R}^2 : y \geqslant e^x\}$ is closed and $K \subset A$. However, $\operatorname{dist}(K, A) = 0$.

Remark 1.28. In addition, the distance function has the following properties:

(1) $x \in \overline{A}$ if and only if $\operatorname{dist}(x, A) = 0$,

(2) $\emptyset \neq A \subset B$ implies $\operatorname{dist}(x, A) \geqslant \operatorname{dist}(x, B)$,

(3) $\operatorname{dist}(x, A) = \operatorname{dist}(x, \overline{A})$ for every $x \in \mathbb{R}^n$ and

(4) $\overline{A} = \overline{B}$ if and only if $\operatorname{dist}(x, A) = \operatorname{dist}(x, B)$ for every $x \in \mathbb{R}^n$.

(Exercise)

Remark 1.29. The distance function can be used to construct a cutoff function, which useful in localization arguments and partitions of unity. Assume that $G \subset \mathbb{R}^n$ is open and $F \subset G$ closed. Then there exist a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ such that

(1) $0 \leqslant f(x) \leqslant 1$ for every $x \in \mathbb{R}^n$,

(2) $f(x) = 1$ for every $x \in F$ and

(3) $f(x) = 0$ for every $x \in \mathbb{R}^n \setminus G$.

Reason. The claim is trivial if F or $\mathbb{R}^n \setminus G$ is empty. Thus we may assume that both sets are nonempty. Define

$$f(x) = \frac{\operatorname{dist}(x, \mathbb{R}^n \setminus G)}{\operatorname{dist}(x, \mathbb{R}^n \setminus G) + \operatorname{dist}(x, F)}.$$

This function has the desired properties. The claim (1) is clear. To prove (2), let $x \in F$. Then $\operatorname{dist}(x, F) = 0$. On the other hand, since $x \in F \subset G$ and G is open, there exits $r > 0$ such that $B(x, r) \subset G$. This implies $\operatorname{dist}(x, \mathbb{R}^n \setminus G) \geqslant r > 0$ and thus $f(x) = 1$. The claim (3) is clear. ■

1.5 Characterizations of measurable sets

In this section we assume that $X = \mathbb{R}^n$ even though most of the results hold true in a more general context. We discuss a useful method to construct σ-algebras. A σ-algebra generated be a collection of sets $\mathscr{E}$ is the smallest σ-algebra containing $\mathscr{E}$. Next we show that that this definition makes sense.

Lemma 1.30. Let $\mathscr{E}$ be a collection of subsets of X. There exists a unique smallest σ-algebra $\mathscr{A}$ containing $\mathscr{E}$, that is, $\mathscr{A}$ is a σ-algebra, $\mathscr{E} \subset \mathscr{A}$, and if $\mathscr{B}$ is any other σ-algebra with $\mathscr{E} \subset \mathscr{B}$, then $\mathscr{A} \subset \mathscr{B}$.

Proof. Let $\mathscr{S}$ be the collection of all σ-algebras $\mathscr{B}$ that contain $\mathscr{E}$ and consider

$$\mathscr{A} = \bigcap_{\mathscr{B}\in\mathscr{S}} \mathscr{B} = \{A \subset X : \text{if } \mathscr{B} \text{ is a } \sigma\text{-algebra with } \mathscr{E} \subset \mathscr{B}, \text{ then } A \in \mathscr{B}\}$$

The collection $\mathscr{A}$ is a σ-algebra, since the intersection of σ-algebras in a σ-algebra. It is easy to verify that $\mathscr{A}$ has the required properties (exercise). □

Definition 1.31. The collection $\mathscr{B}$ of Borel sets is the smallest σ-algebra containing all open subsets of $\mathbb{R}^n$.

Remarks 1.32:

(1) Since any σ-algebra is closed with respect to complements, the collection $\mathscr{B}$ of Borel sets can be also defined as the smallest σ-algebra containing, for example, the closed subsets. In fact $\mathscr{B}$ is generated by open and closed intervals, because every open set is a countable union of open (or closed) intervals by the Lindelöf theorem.

(2) Note that the collection of Borel sets does not only contain open and closed sets, but it also contains, for example, the G_δ-sets which are countable intersections of open sets and F_σ-sets which are countable unions of closed sets. For example, the half-open interval $[0,1)$ is not open nor closed, but both G_δ and F_σ, since it can be expressed as both a countable union of closed sets and a countable intersection of open sets

$$[0,1) = \bigcup_{i=1}^{\infty} \left[0, 1 - \frac{1}{i}\right] = \bigcap_{i=1}^{\infty} \left(-\frac{1}{i}, 1\right).$$

There are also many other Borel sets than open, closed, G_δ and F_σ, see [6, Chapter 11].

Definition 1.33. Let μ^* be an outer measure on $\mathbb{R}^n$.

(1) μ^* is called a Borel outer measure, if all Borel sets are μ^*-measurable.

(2) A Borel outer measure μ^* is called Borel regular, if for every set $A \subset \mathbb{R}^n$ there exists a Borel set B such that $A \subset B$ and $\mu^*(A) = \mu^*(B)$.

(3) μ^* is a Radon outer measure, if μ^* is Borel regular and $\mu^*(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$.

THE MORAL: We shall see that the Lebesgue outer measure is a Radon outer measure. Radon outer measures have many good approximation properties similar to the Lebesgue measure. There is also a natural way to construct Radon outer measures by the Riesz representation theorem, but this will be discussed in the real analysis course.

Remarks 1.34:

(1) In particular, all open and closed sets are measurable for a Borel outer measure. Thus the collection of measurable sets is relatively large.

(2) In general, an outer measure μ^* is called regular, if for every set $A \subset \mathbb{R}^n$ there exists a μ^*-measurable set B such that $A \subset B$ and $\mu^*(A) = \mu^*(B)$. Many natural constructions of outer measures give regular measures, see Remark 1.18.

(3) The local finiteness condition $\mu^*(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$ is equivalent with the condition $\mu(B(x,r)) < \infty$ for every $x \in \mathbb{R}^n$ and $r > 0$. This implies that μ^* is σ-finite, see Definition 1.16.

Examples 1.35:

(1) The Dirac outer measure is a Radon outer measure (exercise).

(2) The counting measure is Borel regular on any metric space X, but it is a Radon outer measure only if every compact subset of X is finite (exercise).

Lemma 1.36. The Lebesgue outer measure m^* is Borel regular.

Proof. We may assume that $m^*(A) < \infty$, for otherwise we may take $B = \mathbb{R}^n$. For every $i = 1,2,\ldots$ there are intervals $I_{j,i}$, $j = 1,2,\ldots$, such that $A \subset \bigcup_{j=1}^{\infty} I_{j,i}$ and

$$m^*(A) \leqslant \sum_{j=1}^{\infty} \operatorname{vol}(I_{j,i}) < m^*(A) + \frac{1}{i}.$$

Denote $B_i = \bigcup_{j=1}^{\infty} I_{j,i}$, $i = 1,2,\ldots$. The set B_i, $i = 1,2,\ldots$, is a Borel set as a countable union of closed intervals. This implies that $B = \bigcap_{i=1}^{\infty} B_i$ is a Borel set. Moreover, since $A \subset B_i$ for every $i = 1,2,\ldots$, we have $A \subset B \subset B_i$. By monotonicity and the definition of the Lebesgue outer measure, this implies

$$m^*(A) \leqslant m^*(B) \leqslant m^*(B_i) = m^*\left(\bigcup_{j=1}^{\infty} I_{j,i}\right) \leqslant \sum_{j=1}^{\infty} \operatorname{vol}(I_{j,i}) < m^*(A) + \frac{1}{i}.$$

By letting $i \to \infty$, we conclude $m^*(A) = m^*(B)$. We shall show later that all Borel sets are Lebesgue measurable, see Lemma 1.50. □

The next results asserts that the Lebesgue outer measure is locally finite.

Lemma 1.37. The Lebesgue outer measure satisfies $m^*(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$.

Proof. Since $K \subset \mathbb{R}^n$ is compact it is closed and bounded. Thus there exists an interval $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$, $a_i, b_i \in \mathbb{R}$, $i = 1,2,\ldots,n$, such that $K \subset I$. By the definition of the definition of the Lebesgue outer measure, this implies

$$m^*(K) \leqslant \operatorname{vol}(I) = (b_1 - a_1)\cdots(b_n - a_n) < \infty.$$ □

Remark 1.38. The Hausdorff measures, defined in Example 1.3 (6), are not necessarily locally finite. For example, the one-dimensional Hausdorff measure $\mathscr{H}^1$ is a Borel regular outer measure but not a Radon outer measure on $\mathbb{R}^2$, because $\mathscr{H}^1(\overline{B(0,1)}) = \infty$ and the closed unit ball $\overline{B(0,1)} = \{x \in \mathbb{R}^2 : |x| \leqslant 1\}$ is a compact subset of $\mathbb{R}^2$. We shall show later that all Borel sets are measurable wit respect to the Hausdorff outer measure, see Remark 1.49.

We discuss an approximation result for a measurable set with respect to a Radon outer measure. In Lemma 1.39 we assume that $\mu^*(\mathbb{R}^n) < \infty$, but Theorem 1.44 shows that the result holds also when $\mu^*(\mathbb{R}^n) = \infty$.

Lemma 1.39. Let μ^* be a Radon outer measure on $\mathbb{R}^n$, $\mu^*(\mathbb{R}^n) < \infty$ and $A \subset \mathbb{R}^n$ a μ^*-measurable set. For every $\varepsilon > 0$ there exists a closed set F and an open set G such that $F \subset A \subset G$, $\mu^*(A \setminus F) < \varepsilon$ and $\mu^*(G \setminus A) < \varepsilon$.

THE MORAL: A measurable set with respect to a Radon outer measure can be approximated by closed sets from inside and open sets from outside in the sense of measure.

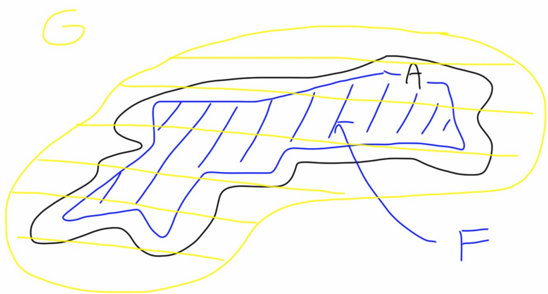

Figure 1.7: Approximation of a measurable set.

Proof. **Step 1:** Let

$$\begin{aligned}\mathscr{F} = \{A \subset \mathbb{R}^n : &A\ \mu^*\text{-measurable, for every } \varepsilon > 0 \text{ there exists a closed } F \subset A \\ &\text{such that } \mu^*(A \setminus F) < \varepsilon \text{ an open } G \supset A \text{ such that } \mu^*(G \setminus A) < \varepsilon\}\end{aligned}$$

be the collection of measurable sets that has the required approximation property.

STRATEGY: We show that $\mathcal{F}$ is a σ-algebra that contains the Borel sets. Borel regularity takes care of the rest.

It is clear that $\emptyset \in \mathcal{F}$ and that $A \in \mathcal{F}$ implies $\mathbb{R}^n \setminus A \in \mathcal{F}$. Let $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$. We show that $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$. Observe that this implies $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$ by de Morgan's law. Let $\varepsilon > 0$. Since $A_i \in \mathcal{F}$ there exist a closed set F_i and an open set G_i such that $F_i \subset A_i \subset G_i$,

$$\mu^*(A_i \setminus F_i) < \frac{\varepsilon}{2^{i+1}} \quad \text{and} \quad \mu^*(G_i \setminus A_i) < \frac{\varepsilon}{2^{i+1}}$$

for every $i = 1, 2, \ldots$ Then

$$\begin{aligned}
\mu^*\left(\bigcap_{i=1}^{\infty} A_i \setminus \bigcap_{i=1}^{\infty} F_i\right) &\leqslant \mu^*\left(\bigcup_{i=1}^{\infty} (A_i \setminus F_i)\right) \quad \text{(monotonicity)} \\
&\leqslant \sum_{i=1}^{\infty} \mu^*(A_i \setminus F_i) \quad \text{(subadditivity)} \\
&\leqslant \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^{i+1}} < \varepsilon.
\end{aligned}$$

Since $\bigcap_{i=1}^{\infty} F_i$ is a closed set, it will do as an approximation from inside. On the other hand, since $\mu^*(\mathbb{R}^n) < \infty$, Theorem 1.20 implies

$$\lim_{k\to\infty} \mu^*\left(\bigcap_{i=1}^{k} G_i \setminus \bigcap_{i=1}^{\infty} A_i\right) = \mu^*\left(\bigcap_{i=1}^{\infty} G_i \setminus \bigcap_{i=1}^{\infty} A_i\right) < \varepsilon.$$

The last inequality is proved as above. Consequently, there exists an index k such that

$$\mu^*\left(\bigcap_{i=1}^{k} G_i \setminus \bigcap_{i=1}^{\infty} A_i\right) < \varepsilon$$

As an intersection of finitely many open sets, $\bigcap_{i=1}^{k} G_i$ is an open set, and it will do as an approximation from outside. This shows that $\mathcal{F}$ is a σ-algebra.

Then we show that $\mathcal{F}$ contains the closed sets. Assume that A is a closed set. $\mu^*(A \setminus A) = 0 < \varepsilon$, so that A itself will do as an approximation from inside. Since A is closed

$$A = \bigcap_{i=1}^{\infty} A_i, \quad \text{where} \quad A_i = \left\{x \in \mathbb{R}^n : \operatorname{dist}(x, A) < \frac{1}{i}\right\}, \; i = 1, 2, \ldots$$

The sets A_i, $i = 1, 2, \ldots$, are open, because $x \mapsto \operatorname{dist}(x, A)$ is continuous, see Lemma 1.25. Since $\mu^*(\mathbb{R}^n) < \infty$ and $A_1 \supset A_2 \supset \ldots$, Theorem 1.20 implies

$$\lim_{i\to\infty} \mu^*(A_i \setminus A) = \mu^*\left(\bigcap_{i=1}^{\infty} A_i \setminus A\right) = \mu^*(\emptyset) = 0,$$

and there exists an index i such that $\mu^*(A_i \setminus A) < \varepsilon$. This A_i will do as an approximation from outside.

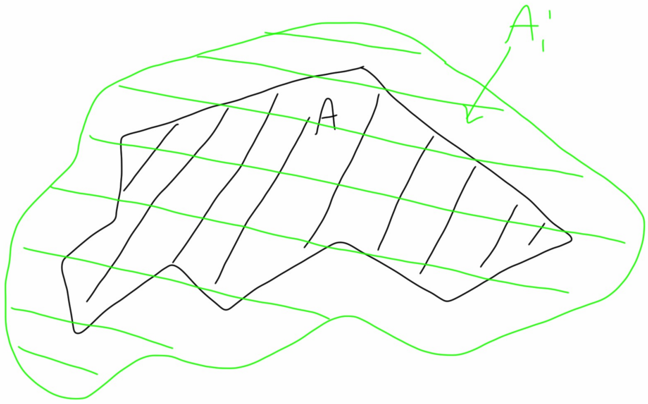

Figure 1.8: A covering of a closed set by distance sets.

Thus $\mathscr{F}$ is σ-algebra containing the closed sets and consequently also the Borel sets. This follows from the fact that the collection of the Borel sets is the smallest σ-algebra with this property. This proves the claim for the Borel sets.

Step 2: Assume then A is a general μ^*-measurable set. By Borel regularity there exists a Borel set $B_1 \supset A$ with $\mu^*(B_1) = \mu^*(A)$ and a Borel set B_2 with $\mathbb{R}^n \setminus A \subset \mathbb{R}^n \setminus B_2$ and $\mu^*(\mathbb{R}^n \setminus B_2) = \mu^*(\mathbb{R}^n \setminus A)$. By Step 1 of the proof there exist a closed set F an open set G such that $F \subset B_2 \subset A \subset B_1 \subset G$,

$$\mu^*(G \setminus B_1) < \varepsilon \quad \text{and} \quad \mu^*(B_2 \setminus F) < \varepsilon.$$

It follows that

$$\begin{aligned}
\mu^*(G \setminus A) &\leqslant \mu^*(G \setminus B_1) + \mu^*(B_1 \setminus A) \quad \text{(subadditivity, even = holds)} \\
&< \varepsilon + \mu^*(B_1) - \mu^*(A) \quad (\mu^*(B_1 \setminus A) = \mu^*(B_1) - \mu^*(A), \mu^*(A) < \infty) \\
&= \varepsilon \quad (\mu^*(B_1) - \mu^*(A) = 0)
\end{aligned}$$

and

$$\begin{aligned}
\mu^*(A \setminus F) &\leqslant \mu^*(A \setminus B_2) + \mu^*(B_2 \setminus F) \\
&\leqslant \mu^*\Big((\mathbb{R}^n \setminus B_2) \setminus (\mathbb{R}^n \setminus A)\Big) + \varepsilon \\
&= \mu^*(\mathbb{R}^n \setminus B_2) - \mu^*(\mathbb{R}^n \setminus A) + \varepsilon \quad (\mathbb{R}^n \setminus A \text{ measurable}, \ \mu^*(\mathbb{R}^n \setminus A) < \infty) \\
&= \varepsilon. \quad (\mu^*(\mathbb{R}^n \setminus B_2) - \mu^*(\mathbb{R}^n \setminus A) = 0, \ \mu^*(\mathbb{R}^n) < \infty)
\end{aligned}$$

$\square$

Definition 1.40. Let μ^* be an outer measure on $\mathbb{R}^n$ and E an arbitrary subset of $\mathbb{R}^n$. Then the restriction of μ^* to E is defined to be

$$(\mu^* \lfloor E)(A) = \mu^*(A \cap E)$$

for every $A \subset \mathbb{R}^n$.

T H E M O R A L : The restriction of a measure is a useful tool to make an outer measure finite by considering a restriction to a set with finite measure.

Figure 1.9: The restriction of a measure.

Remarks 1.41:

(1) $\mu^* \lfloor E$ is an outer measure (exercise).

(2) Any μ^*-measurable set is also $\mu^* \lfloor E$-measurable (exercise). This holds for all sets $E \subset \mathbb{R}^n$. In particular, the set E does not have to be μ^*-measurable. Note that not all $\mu^* \lfloor E$-measurable sets need not be μ^*-measurable.

(3) It is useful to consider restrictions of the Hausdorff measures, defined in Example 1.3, to sets with a lower Hausdorff dimension than n. Consider, for example, the one-dimensional Hausdorff measure $\mathscr{H}^1$ on $\mathbb{R}^2$. By Remark 1.38 we have $\mathscr{H}^1(\mathbb{R}^2) = \infty$, but $(\mathscr{H}^1 \lfloor \partial B(0,1))(\mathbb{R}^2) < \infty$.

Lemma 1.42. Let μ^* be a Borel regular outer measure on $\mathbb{R}^n$. Assume that $E \subset \mathbb{R}^n$ is μ^*-measurable and $\mu^*(E) < \infty$. Then $\mu^* \lfloor E$ is a Radon outer measure.

Remarks 1.43:

(1) The assumption $\mu^*(E) < \infty$ cannot be removed, see Remark 1.38.

(2) If E is a Borel set, then $\mu^* \lfloor E$ is Borel regular even if $\mu^*(E) = \infty$ (exercise).

Proof. Let $\nu = \mu^* \lfloor E$. Since every μ^*-measurable set is ν-measurable, ν is a Borel outer measure. If $K \subset \mathbb{R}^n$ is compact, then

$$\nu(K) = \mu^*(K \cap E) \leqslant \mu^*(E) < \infty.$$

CLAIM: ν is Borel regular.

Since μ^* is Borel regular, there exists a Borel set B_1 such that $E \subset B_1$ and $\mu^*(B_1) = \mu^*(E)$. Then

$$\begin{aligned} \mu^*(B_1) &= \mu^*(B_1 \cap E) + \mu^*(B_1 \setminus E) \quad (E \text{ is } \mu^*\text{-measurable}) \\ &= \mu^*(E) + \mu^*(B_1 \setminus E) \quad (E \subset B_1) \end{aligned}$$

Since $\mu^*(E) < \infty$, we have $\mu^*(B_1 \setminus E) = \mu^*(B_1) - \mu^*(E) = 0$.

Let $A \subset \mathbb{R}^n$. Since μ^* is Borel regular, there exists a Borel set B_2 such that $B_1 \cap A \subset B_2$ and $\mu^*(B_2) = \mu^*(B_1 \cap A)$. Then $A \subset B_2 \cup (\mathbb{R}^n \setminus B_1) = C$ and C is a Borel set as a union of two Borel sets. We have

$$\begin{aligned} (\mu^* \lfloor E)(C) &= \mu^*(C \cap E) \leqslant \mu^*(B_1 \cap C) \qquad (E \subset B_1) \\ &= \mu^*(B_1 \cap B_2) \quad (B_1 \cap C = B_1 \cap (B_2 \cup (\mathbb{R}^n \setminus B_1)) = B_1 \cap B_2) \\ &\leqslant \mu^*(B_2) = \mu^*(B_1 \cap A) \\ &= \mu^*((B_1 \cap A) \cap E) + \mu^*((B_1 \cap A) \setminus E) \quad (E \text{ is } \mu^*\text{-measurable}) \\ &\leqslant \mu^*(E \cap A) + \mu^*(B_1 \setminus E) \quad (\text{monotonicity}) \\ &= (\mu^* \lfloor E)(A). \quad (\mu^*(B_1 \setminus E) = 0) \end{aligned}$$

On the other hand, $A \subset C$ implies $(\mu^* \lfloor E)(A) \leqslant (\mu^* \lfloor E)(C)$. Consequently $(\mu^* \lfloor E)(A) = (\mu^* \lfloor E)(C)$ and $\mu^* \lfloor E$ is Borel regular. □

Next we discuss the first characterization of measurable set with respect to a Radon outer measure.

Theorem 1.44. Let μ^* be a Radon outer measure on $\mathbb{R}^n$. Then the following conditions are equivalent:

(1) $A \subset \mathbb{R}^n$ is μ^*-measurable,

(2) for every $\varepsilon > 0$ there exists a closed set F and an open set G such that $F \subset A \subset G$, $\mu^*(A \setminus F) < \varepsilon$ and $\mu^*(G \setminus A) < \varepsilon$.

THE MORAL: This is a topological characterization of a measurable set through an approximation property. A set is measurable for a Radon outer measure if and only if it can be approximated by closed sets from inside and open sets from outside in the sense of measure. Observe that the original Carathéodory criterion for measurablity in Definition 1.4 depends only on the outer measure and there is no reference to open or closed sets.

Proof. $\boxed{(1) \Rightarrow (2)}$ Let $\nu_i = \mu^* \lfloor B(0,i)$, with $B(0,i) = \{x \in \mathbb{R}^n : |x| < i\}$, $i = 1,2,\dots$ By Lemma 1.42, ν_i is a Radon outer measure and $\nu_i(\mathbb{R}^n) \leqslant \mu^*(\overline{B(0,i)}) < \infty$ for every $i = 1,2,\dots$ Since A is μ^*-measurable, A is also ν_i-measurable.

By Lemma 1.39, there exists an open set $G_i \supset A$ such that

$$\nu_i(G_i \setminus A) < \frac{\varepsilon}{2^{i+1}},$$

for every $i = 1,2,\dots$. Let

$$G = \bigcup_{i=1}^{\infty} \Big(G_i \cap B(0,i)\Big)$$

As a union of open sets, the set G is open and $G \supset A$. Moreover,

$$\begin{aligned}
\mu^*(G \setminus A) &= \mu^*\left(\left(\bigcup_{i=1}^{\infty}\Big(G_i \cap B(0,i)\Big)\right) \setminus A\right) \\
&= \mu^*\left(\bigcup_{i=1}^{\infty}\Big((G_i \setminus A) \cap B(0,i)\Big)\right) \\
&\leqslant \sum_{i=1}^{\infty} \mu^*\Big((G_i \setminus A) \cap B(0,i)\Big) \\
&\leqslant \sum_{i=1}^{\infty} \nu_i(G_i \setminus A) \leqslant \sum_{i=1}^{\infty} \frac{\varepsilon}{2^{i+1}} < \varepsilon.
\end{aligned}$$

Thus G will do as an approximation from outside.

By considering the complements, there exists a closed set F such that $\mathbb{R}^n \setminus F \supset \mathbb{R}^n \setminus A$ and

$$\mu^*(A \setminus F) = \mu^*((\mathbb{R}^n \setminus F) \setminus (\mathbb{R}^n \setminus A)) < \varepsilon.$$

The set F is closed and will do as an approximation from inside.

$\boxed{(2) \Rightarrow (1)}$ For every $i = 1,2,\dots$ there exists a closed set $F_i \subset A$ such that $\mu^*(A \setminus F_i) < \frac{1}{i}$. Then $F = \bigcup_{i=1}^{\infty} F_i$ is a Borel set (not necessarily closed) and $F \subset A$. Moreover,

$$\mu^*(A \setminus F) = \mu^*\left(A \setminus \bigcup_{i=1}^{\infty} F_i\right) = \mu^*\left(\bigcap_{i=1}^{\infty}(A \setminus F_i)\right) \leqslant \mu^*(A \setminus F_i) < \frac{1}{i}$$

for every $i = 1,2,\dots$ This implies

$$0 \leqslant \lim_{i\to\infty} \mu^*(A \setminus F) \leqslant \lim_{i\to\infty} \frac{1}{i} = 0$$

and consequently $\mu^*(A \setminus F) = 0$. Observe that $A = F \cup (A \setminus F)$, where F is a Borel set and hence μ^*-measurable. On the other hand, $\mu^*(A \setminus F) = 0$ so that $A \setminus F$ is μ^*-measurable. The set A is μ^*-measurable as a union of two measurable sets. $\square$

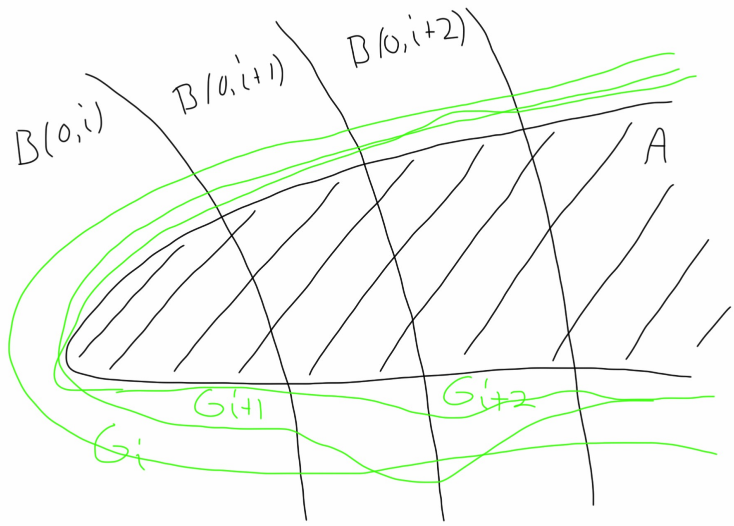

Figure 1.10: Approximation by open sets in measure.

Remark 1.45. We can see from the proof that, for a Radon outer measure, an arbitrary measurable set differs from a Borel set only by a set of measure zero. A set A is μ^*-measurable if and only if $A = F \cup N$, where F is a Borel set and $\mu^*(N) = 0$. Moreover, the set F can be chosen to be a countable union of closed sets, that is, an F_σ set. On the other hand, a set A is μ^*-measurable if and only if $A = G \setminus N$, where g is a Borel set and $\mu^*(N) = 0$. Moreover, we the set G can be chosen to be a countable intersection of open sets, that is, a G_δ set.

Corollary 1.46. Let μ^* be a Radon outer measure on $\mathbb{R}^n$.

(1) (Outer measure) For every set $A \subset \mathbb{R}^n$,

$$\mu^*(A) = \inf\{\mu^*(G) : A \subset G,\ G \text{ open}\}.$$

(2) (Inner measure) For every μ^*-measurable set $A \subset \mathbb{R}^n$,

$$\mu^*(A) = \sup\{\mu^*(K) : K \subset A,\ K \text{ compact}\}.$$

THE MORAL: The inner and outer measures coincide for a measurable set. In this case, the measure can be determined by compact sets from inside or open sets from outside.

Proof. $\boxed{(1)}$ If $\mu^*(A) = \infty$, the claim is clear. Hence we may assume that $\mu^*(A) < \infty$.

Step 1: Assume that A is a Borel set and let $\varepsilon > 0$. Since μ^* is a Borel outer measure, the set A is μ^*-measurable. By Theorem 1.44, there exists an open set $G \supset A$ such that $\mu^*(G \setminus A) < \varepsilon$. Moreover,

$$\begin{aligned}\mu^*(G) &= \mu^*(G \cap A) + \mu^*(G \setminus A) \quad (A \text{ is } \mu^*\text{-measurable})\\ &= \mu^*(A) + \mu^*(G \setminus A) < \mu^*(A) + \varepsilon. \quad (A \subset G)\end{aligned}$$

This implies the claim.

Step 2: Assume then that $A \subset \mathbb{R}^n$ is an arbitrary set. Since μ^* is Borel regular, there exists a Borel set $B \supset A$ with $\mu^*(B) = \mu^*(A)$. It follows that

$$\begin{aligned}\mu^*(A) = \mu^*(B) &= \inf\{\mu^*(G) : B \subset G,\ G \text{ open}\} \quad (\text{Step 1})\\ &\geqslant \inf\{\mu^*(G) : A \subset G,\ G \text{ open}\}. \quad (A \subset B)\end{aligned}$$

On the other hand, by monotonicity,

$$\mu^*(A) \leqslant \inf\{\mu^*(G) : A \subset G,\ G \text{ open}\}$$

and, consequently, the equality holds.

$\boxed{(2)}$ Assume first that $\mu^*(A) < \infty$ and let $\varepsilon > 0$. By Theorem 1.44, there exists a closed set $F \subset A$ such that $\mu^*(A \setminus F) < \varepsilon$. Since F is μ^*-measurable and $\mu^*(A) < \infty$, we have

$$\mu^*(A) - \mu^*(F) = \mu^*(A \setminus F) < \varepsilon$$

and thus $\mu^*(F) > \mu^*(A) - \varepsilon$. This implies that

$$\mu^*(A) = \sup\{\mu^*(F) : F \subset A,\ F \text{ closed}\}$$

Then we consider the case $\mu^*(A) = \infty$. Let $B_i = \{x \in \mathbb{R}^n : i-1 \leqslant |x| < i\}$, $i = 1,2,\ldots$. Then $A = \bigcup_{i=1}^{\infty}(A \cap B_i)$ and by Theorem 1.10

$$\sum_{i=1}^{\infty} \mu^*(A \cap B_i) = \mu^*(A) = \infty$$

because the sets $A \cap B_i$, $i = 1,2,\ldots$, are pairwise disjoint and μ^*-measurable. Since μ^* is a Radon outer measure, $\mu^*(A \cap B_i) \leqslant \mu^*(\overline{B_i}) < \infty$. By the beginning of the proof, there exists a closed set $F_i \subset A \cap B_i$ such that

$$\mu^*(F_i) > \mu^*(A \cap B_i) - \frac{1}{2^i}$$

for every $i = 1,2,\ldots$ Clearly $\bigcup_{i=1}^{\infty} F_i \subset A$ and

$$\begin{aligned}\lim_{k\to\infty} \mu^*\left(\bigcup_{i=1}^{k} F_i\right) &= \mu^*\left(\bigcup_{i=1}^{\infty} F_i\right) \qquad (\text{Theorem } 1.20)\\ &= \sum_{i=1}^{\infty} \mu^*(F_i) \quad (F_i \text{ disjoint } (F_i \subset A \cap B_i), \text{ Theorem } 1.10)\\ &\geqslant \sum_{i=1}^{\infty}\left(\mu^*(A \cap B_i) - \frac{1}{2^i}\right) = \infty.\end{aligned}$$

The set $F = \bigcup_{i=1}^{k} F_i$ is closed as a union of finitely many closed sets and hence

$$\mu^*(A) = \sup\{\mu^*(F) : F \subset A,\ F \text{ closed}\} = \infty.$$

Finally we pass over to compact sets. Assume that F is closed. Then the sets $F \cap \overline{B(0,i)}$, $i = 1,2,\dots$, are closed and bounded and hence compact. By Theorem 1.20,

$$\mu^*(F) = \mu^*\left(\bigcup_{i=1}^{\infty}(F \cap \overline{B(0,i)})\right) = \lim_{i\to\infty} \mu^*(F \cap \overline{B(0,i)})$$

and consequently

$$\sup\{\mu^*(K) : K \subset A,\ K \text{ compact}\} = \sup\{\mu^*(F) : F \subset A,\ F \text{ closed}\}. \qquad \square$$

Remark 1.47. Let μ^* be a Radon outer measure on $\mathbb{R}^n$. There is a delicate issue related to the approximation by open sets. By Corollary 1.46, for every $\varepsilon > 0$, there exists an open set $G \supset A$ with $\mu^*(G) \leqslant \mu^*(A) + \varepsilon$ for every set $A \subset \mathbb{R}^n$. On the other hand, by Theorem 1.44, for every $\varepsilon > 0$, there exists an open set G such that $A \subset G$ and $\mu^*(G \setminus A) \leqslant \varepsilon$ for every μ^*-measurable set $A \subset \mathbb{R}^n$. Observe that these are two different claims, if A does not satisfy the measurability condition in Definition 1.4.

1.6 Metric outer measures

Next we give a useful method to show that Borel sets are measurable.

Theorem 1.48. Let μ^* be an outer measure on $\mathbb{R}^n$. If μ^* is a metric outer measure, that is,

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

for every $A, B \subset \mathbb{R}^n$ with $\operatorname{dist}(A,B) > 0$, then μ^* is a Borel outer measure.

THE MORAL: If an outer measure is additive on separated sets, then all Borel sets are measurable. This is a practical way to show that Borel sets are measurable. This means that the very useful properties of an outer measure on measurable sets are available for a large class of sets.

Proof. We shall show that every closed set $F \subset \mathbb{R}^n$ is μ^*-measurable. It is enough to show that

$$\mu^*(E) \geqslant \mu^*(E \cap F) + \mu^*(E \setminus F)$$

for every $E \subset \mathbb{R}^n$. If $\mu^*(E) = \infty$, there is nothing to prove. Hence we may assume that $\mu^*(E) < \infty$. The set $G = \mathbb{R}^n \setminus F$ is open. We separate the set $A = E \setminus F$ from F by considering the sets

$$A_i = \left\{x \in A : \operatorname{dist}(x,F) \geqslant \frac{1}{i}\right\}, \quad i = 1,2,\dots,$$

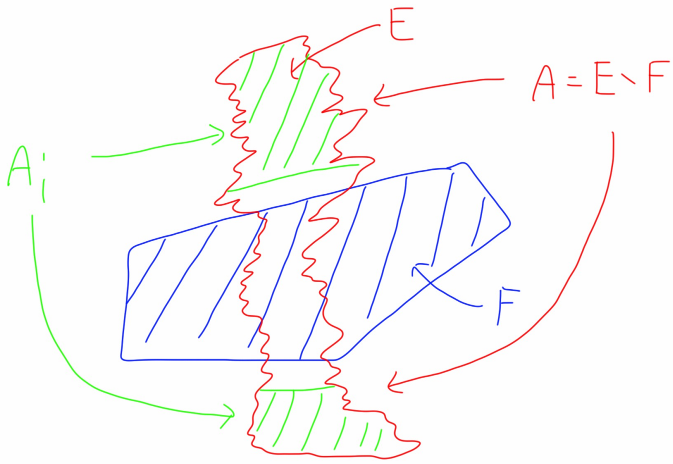

Figure 1.11: Exhaustion by distance sets.

Then $\operatorname{dist}(A_i, F) \geqslant \frac{1}{i}$ for every i= $1,2,\dots$ and $A = \bigcup_{i=1}^{\infty} A_i$, see Lemma 1.26.

Claim: $\lim_{i\to\infty} \mu^*(A_i) = \mu^*(A)$.

Reason. $A_i \subset A_{i+1}$ and $A_i \subset A$, $i = 1,2,\dots$, imply

$$\lim_{i\to\infty} \mu^*(A_i) \leqslant \mu^*(A).$$

Then we prove the reverse inequality. Let

$$B_i = A_{i+1} \setminus A_i = \left\{ x \in A : \frac{1}{i+1} \leqslant \operatorname{dist}(x,F) < \frac{1}{i} \right\}, \quad i = 1,2,\dots$$

Since $A_i \subset A_{i+1}$ for $i = 1,2,\dots$, we have $A = \bigcup_{i=1}^{\infty} A_i = A_i \cup \bigcup_{j=i}^{\infty} B_j$ and by subadditivity

$$\mu^*(A) \leqslant \mu^*(A_i) + \sum_{j=i}^{\infty} \mu^*(B_j).$$

It follows that

$$\mu^*(A) \leqslant \lim_{i\to\infty} \mu^*(A_i) + \lim_{i\to\infty} \sum_{j=i}^{\infty} \mu^*(B_j),$$

where

$$\lim_{i\to\infty} \sum_{j=i}^{\infty} \mu^*(B_j) = 0, \text{ provided } \sum_{j=1}^{\infty} \mu^*(B_j) < \infty.$$

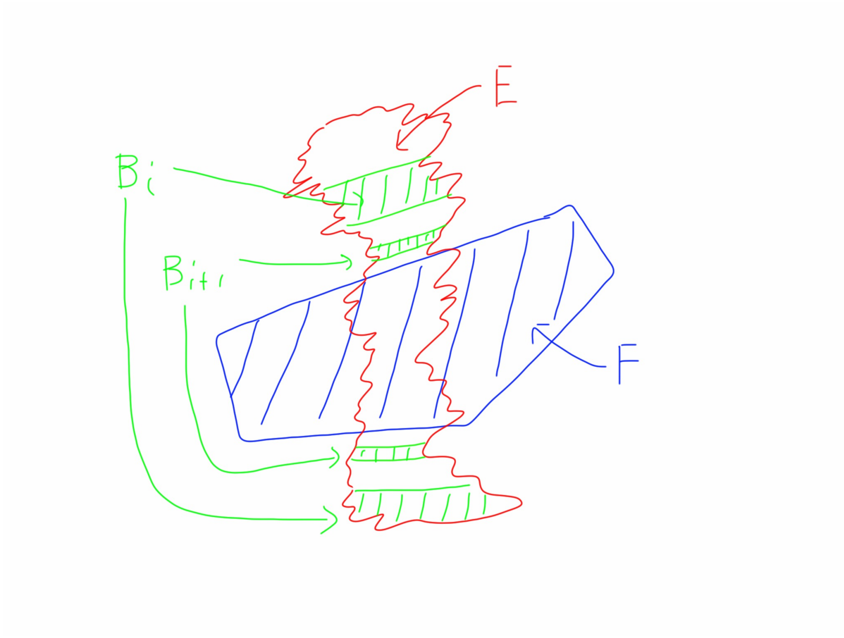

Figure 1.12: Exhaustion by separated distance sets.

By the construction $\operatorname{dist}(B_j, B_l) > 0$ whenever $l \geqslant j+2$. By the assumption we have

$$\sum_{j=1}^{k} \mu^*(B_{2j}) = \mu^*\left(\bigcup_{j=1}^{k} B_{2j}\right) \leqslant \mu^*(E) < \infty$$

and

$$\sum_{j=0}^{k} \mu^*(B_{2j+1}) = \mu^*\left(\bigcup_{j=0}^{k} B_{2j+1}\right) \leqslant \mu^*(E) < \infty.$$

These together imply

$$\sum_{j=1}^{\infty} \mu^*(B_j) = \lim_{k\to\infty}\left(\sum_{j=1}^{k} \mu^*(B_{2j}) + \sum_{j=0}^{k} \mu^*(B_{2j+1})\right) \leqslant 2\mu^*(E) < \infty.$$

Thus

$$\mu^*(A) \leqslant \lim_{i\to\infty} \mu^*(A_i)$$

and consequently

$$\lim_{i\to\infty} \mu^*(A_i) = \mu^*(A).$$

■

Finally

$$\begin{aligned}
\mu^*(E\cap F) + \mu^*(E\setminus F) &= \mu^*(E\cap F) + \mu^*(A) \quad (A = E\setminus F)\\
&= \lim_{i\to\infty}\Big(\mu^*(E\cap F) + \mu^*(A_i)\Big) \quad \text{(above)}\\
&= \lim_{i\to\infty} \mu^*\Big((E\cap F)\cup A_i\Big) \quad (\operatorname{dist}(A_i, F) > 0)\\
&\leqslant \mu^*(E). \quad \text{(monotonicity)}
\end{aligned}$$

□

Remarks 1.49:

(1) The converse holds as well, so that the previous theorem gives a characterization for a Borel outer measure. Observe, that there may be also other measurable sets than Borel sets, because an arbitrary measurable set for a Radon measure can be represented as a union of Borel set and a set of measure zero, see Remark 1.45.

(2) The Carathéodory construction in Example 1.3 (8) always gives a metric outer measure. In particular, all Borel sets are measurable. Moreover, if the members of covering family in the construction are Borel sets, then the measure is Borel regular (exercise). Thus many natural constructions give a Borel regular outer measure.

(3) The Hausdorff measure, defined in Example 1.3 (6), is a metric outer measures (exercise). Thus all Borel sets are measurable with respect to a Hausdorff measure. See also [2, Section 3.8], [4, Chapter 2], [8, Chapter 19], [11, Chapter 7] and [16, Chapter 7].

Lemma 1.50. The Lebesgue outer measure m^* is a metric outer measure.

THE MORAL: Theorem 1.48 implies that the Lebesgue outer measure m^* is a Borel outer measure. Thus all Borel sets are m^*-measurable, in particular closed and open sets, are Lebesgue measurable. By Lemmas 1.36 and 1.37 we can conclude that m^* is a Radon outer measure.

Proof. Let $A,B \subset \mathbb{R}^n$ with $\operatorname{dist}(A,B) > 0$. Subadditivity implies that $m^*(A \cup B) \leqslant m^*(A) + m^*(B)$. Hence it is enough to prove the reverse inequality. For every $\varepsilon > 0$ there are intervals I_i, $i = 1,2,\dots$, such that $A \cup B \subset \bigcup_{i=1}^{\infty} I_i$ and

$$\sum_{i=1}^{\infty} \operatorname{vol}(I_i) < m^*(A \cup B) + \varepsilon.$$

By subdividing each I_i into smaller intervals, we may assume that $\operatorname{diam}(I_i) < \operatorname{dist}(A,B)$ for every $i = 1,2,\dots$ In this case every interval I_i intersects at most one of the sets A and B.

THE MORAL: We can assume that the diameter of the intervals in the definition of the Lebesgue measure is as small as we want.

We consider two subfamilies I'_i and I''_i, $i = 1,2,\dots$, where the intervals of the first have a nonempty intersection with A and the intervals of the second have a nonempty intersection with B. Note that there may be intervals that do not intersect $A \cup B$, but this is not a problem. Thus

$$m^*(A) + m^*(B) \leqslant \sum_{i=1}^{\infty} \operatorname{vol}(I'_i) + \sum_{i=1}^{\infty} \operatorname{vol}(I''_i) \leqslant \sum_{i=1}^{\infty} \operatorname{vol}(I_i) < m^*(A \cup B) + \varepsilon. \qquad \square$$

By letting $\varepsilon \to 0$, we obtain $m^*(A \cup B) = m^*(A) + m^*(B)$.

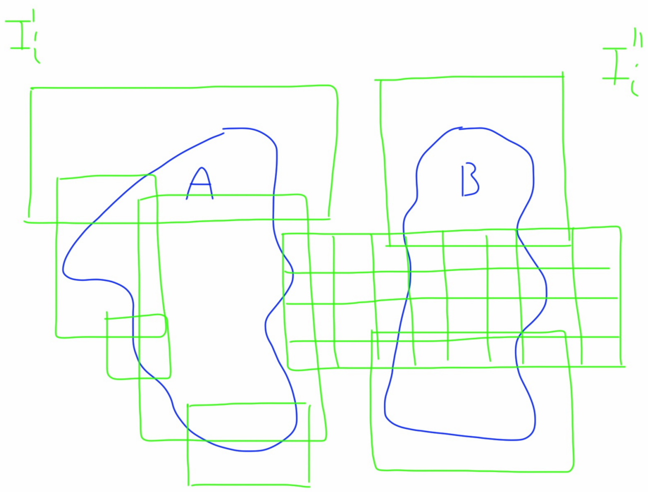

Figure 1.13: Lebesgue measure is a metric measure.

1.7 Lebesgue measure revisited

We have already discussed the definition of the Lebesgue outer measure in Example 1.3 (5). Recall that the Lebesgue outer measure of an arbitrary set $A \subset \mathbb{R}^n$ is

$$m^*(A) = \inf\left\{\sum_{i=1}^{\infty} \operatorname{vol}(I_i) : A \subset \bigcup_{i=1}^{\infty} I_i\right\},$$

where the infimum is taken over all coverings of A by countably many closed intervals I_i, $i = 1, 2, \dots$. We discuss examples and properties that are characteristic for the Lebesgue outer measure.

Remarks 1.51:

(1) We cannot upgrade countable subadditivity of the Lebesgue outer measure to uncountable subadditivity. For example, $\mathbb{R}^n$ is an uncountable union of points, each of which has Lebesgue outer measure zero, but $\mathbb{R}^n$ has infinite Lebesgue outer measure.

(2) If we consider coverings with finitely many intervals, we obtain the Jordan outer measure defined as

$$m^{*,J}(A) = \inf\left\{\sum_{i=1}^{k} \operatorname{vol}(I_i) : A \subset \bigcup_{i=1}^{k} I_i,\ k = 1, 2, \dots\right\},$$

where $A \subset \mathbb{R}^n$ is a bounded set. The Jordan outer measure will not be an outer measure, since is only finitely subadditive instead of countably subadditive. It has the property $J^*(A) = J^*(\overline{A})$ for every bounded $A \subset \mathbb{R}^n$. We can define the corresponding Jordan inner measure by

$$m_{*,J}(A) = \sup\left\{\sum_{i=1}^{k} \operatorname{vol}(I_i) : A \supset \bigcup_{i=1}^{k} I_i,\ k = 1,2,\dots\right\}$$

and say that a bounded set $A \subset \mathbb{R}^n$ is Jordan measurable if the inner and outer Jordan measures coincide. It can be shown that a bounded set $A \subset \mathbb{R}^n$ is Jordan measurable if and only if the Jordan outer measure of ∂A is zero. For example,

$$m^{*,J}(\mathbb{Q} \cap [0,1]) = 1 \quad \text{and} \quad m_{*,J}(\mathbb{Q} \cap [0,1]) = 0,$$

while $m^*(\mathbb{Q} \cap [0,1])) = 0$, since $\mathbb{Q} \cap [0,1]$ is a countable set. In particular, $Q \cap [0,1]$ is Lebesgue measurable but not Jordan measurable. This example also shows that the Jordan outer measure is not countably additive. See [12] for more on the Jordan outer measure.

(3) The closed intervals in the definition of the Lebesgue outer measure can be replaced by open intervals or cubes. Cubes are intervals whose side lenghts are equal, that is $b_1 - a_1 = \dots = b_n - a_n$. Even balls will do, but this is more subtle (exercise).

Next we show that the Lebesgue outer measure of a closed interval is equal to its volume. Note that this is not obvious from the definition. This means that the definition of the Lebesgue outer measure is consistent for closed intervals.

Lemma 1.52. Let $I \subset \mathbb{R}^n$ be a closed interval. Then $m^*(I) = \operatorname{vol}(I)$.

Proof. It is clear that $m^*(I) \leqslant \operatorname{vol}(I)$, since the interval I itself is an admissible covering the definition of the Lebesgue outer measure. Hence it remains to prove that $\operatorname{vol}(I) \leqslant m^*(I)$. For every $\varepsilon > 0$ there are intervals I_j, $j = 1,2,\dots$, such that $I \subset \bigcup_{j=1}^{\infty} I_j$ and

$$\sum_{j=1}^{\infty} \operatorname{vol}(I_j) < m^*(I) + \varepsilon.$$

For every $j = 1,2,\dots$ there is an open interval J_j such that $I_j \subset J_j$ and

$$\operatorname{vol}(J_j) \leqslant \operatorname{vol}(I_j) + \frac{\varepsilon}{2^j}.$$

This implies

$$\sum_{j=1}^{\infty} \operatorname{vol}(J_j) \leqslant \sum_{j=1}^{\infty} \left(\operatorname{vol}(I_j) + \frac{\varepsilon}{2^j}\right) = \sum_{j=1}^{\infty} \operatorname{vol}(I_j) + \sum_{j=1}^{\infty} \frac{\varepsilon}{2^j} = \sum_{j=1}^{\infty} \operatorname{vol}(I_j) + \varepsilon.$$

The collection of intervals J_j, $j = 1,2,\dots$, is an open covering of the compact set I and there exist a finite subcovering J_j, $j = 1,2,\dots,k$.

We split I into finitely many closed subintervals $K_1,\dots,K_m$ such that $I = \bigcup_{i=1}^{m} K_i$, the interiors of $K_1,\dots,K_m$ are pairwise disjoint and $K_i \subset J_{j_i}$ for some j_i

with $j_i \leqslant k$. Note that more than one K_i may belong to the same J_{j_i}. This can be done by considering the grid obtained by extending indefinitely the sides of all intervals $J_1,\dots,J_k$. Then

$$\operatorname{vol}(I)=\sum_{i=1}^{m}\operatorname{vol}(K_i)$$

(exercise), $\bigcup_{i=1}^{m} K_i \subset \bigcup_{j=1}^{k} J_i$ and

$$\begin{aligned}\operatorname{vol}(I)=\sum_{i=1}^{m}\operatorname{vol}(K_i) &\leqslant \sum_{j=1}^{k}\sum_{\{i:K_i\subset J_j\}}\operatorname{vol}(K_i)\leqslant\sum_{j=1}^{k}\operatorname{vol}(J_j)\\ &\leqslant\sum_{j=1}^{\infty}\operatorname{vol}(J_j)\leqslant\sum_{j=1}^{\infty}\operatorname{vol}(I_j)+\varepsilon\leqslant m^*(I)+2\varepsilon.\end{aligned}$$

The claim follows by letting $\varepsilon\to 0$. □

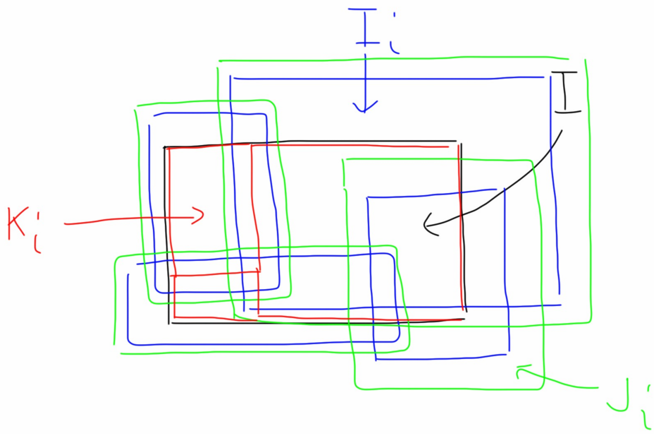

Figure 1.14: Lebesgue measure of an interval.

Remark 1.53. The assertion $m^*(I)=\operatorname{vol}(I)$ in Lemma 1.52 holds for an open interval $I\subset\mathbb{R}^n$ as well. We discuss three ways to prove this claim.

$\boxed{(1)}$ Since I is covered by its closure $\overline{I}$, we have $m^*(I)\leqslant\operatorname{vol}(\overline{I})=\operatorname{vol}(I)$. To prove the reverse inequality, let $\varepsilon>0$ and let J be a closed interval contained in I with $\operatorname{vol}(I)\leqslant\operatorname{vol}(J)+\varepsilon$. By monotonicity $m^*(J)\leqslant m^*(I)$ and by Lemma 1.52 we obtain

$$\operatorname{vol}(I)\leqslant\operatorname{vol}(J)+\varepsilon=m^*(J)+\varepsilon\leqslant m^*(I)+\varepsilon.$$

By letting $\varepsilon\to 0$, we have $\operatorname{vol}(I)\leqslant m^*(I)$ and consequently $\operatorname{vol}(I)\leqslant m^*(I)$.

$\boxed{(2)}$ In open interval $I \subset \mathbb{R}^n$ can be written as a union of an increasing sequence of closed intervals I_i, $i = 1, 2, \dots$. By Theorem 1.20 we have

$$m^*(I) = m^*\left(\bigcup_{i=1}^{\infty} I_i\right) = \lim_{i\to\infty} m^*(I_i) = \operatorname{vol}(I).$$

$\boxed{(3)}$ The claim also follows from the fact that $m^*(\partial I) = 0$, since open and closed intervals are Lebesgue measurable (exercise).

Next we discuss the Lebesgue outer measure of countable unions of intervals. We say that closed intervals I_i, $i = 1, 2, \dots$, are almost disjoint if their interiors are pairwise disjoint. Thus almost disjoint intervals may touch only on the boundaries.

Lemma 1.54. If $I_i \subset \mathbb{R}^n$, $i = 1, 2, \dots$ are almost disjoint closed intervals, then

$$m^*\left(\bigcup_{i=1}^{\infty} I_i\right) = \sum_{i=1}^{\infty} m^*(I_i).$$

T H E   M O R A L : If a set can be decomposed into almost disjoint intervals, its Lebesgue outer measure equals the sum of the volumes of the intervals. In principle, this gives us a method to compute the Lebesgue outer measure of a nice set. Remark 1.55 below shows that every open set has this property.

Proof. Since the intervals I_i, $i = 1, 2, \dots$, cover $\bigcup_{i=1}^{\infty} I_i$, by the definition of the Lebesgue outer measure

$$m^*\left(\bigcup_{i=1}^{\infty} I_i\right) \leqslant \sum_{i=1}^{\infty} \operatorname{vol}(I_i).$$

Then we prove the reverse inequality. For every $i = 1, 2, \dots$, let J_i be a closed interval contained in I_i with $J_i \cap \partial I_i = \emptyset$ and

$$\operatorname{vol}(I_i) \leqslant \operatorname{vol}(J_i) + \frac{\varepsilon}{2^i}.$$

For every $k = 1, 2, \dots$, the intervals $J_1, \dots, J_k$ are pairwise disjoint compact sets and thus $\operatorname{dist}(J_i, J_j) > 0$ for $i \neq j$. Since the Lebesgue outer measure is a metric outer measure, see Lemma 1.50, we have

$$\begin{aligned} m^*\left(\bigcup_{i=1}^{k} J_i\right) &= \sum_{i=1}^{k} m^*(J_i) = \sum_{i=1}^{k} \operatorname{vol}(J_i) \\ &\geqslant \sum_{i=1}^{k} \left(\operatorname{vol}(I_i) - \frac{\varepsilon}{2^i}\right) = \sum_{i=1}^{k} \operatorname{vol}(I_i) - \varepsilon \end{aligned}$$

Here we also used Lemma 1.52. Since $\bigcup_{i=1}^{k} J_i \subset \bigcup_{i=1}^{\infty} I_i$, by monotonicity we have

$$m^*\left(\bigcup_{i=1}^{\infty} I_i\right) \geqslant m^*\left(\bigcup_{i=1}^{k} J_i\right) \geqslant \sum_{i=1}^{k} \operatorname{vol}(I_i) - \varepsilon$$

for every $k = 1,2,\dots$. By letting $k \to \infty$, we obtain

$$\sum_{i=1}^{\infty} \operatorname{vol}(I_i) = \lim_{k\to\infty} \sum_{i=1}^{k} \operatorname{vol}(I_i) \leqslant m^*\left(\bigcup_{i=1}^{\infty} I_i\right) + \varepsilon.$$

Finally, by letting $\varepsilon \to 0$, we have

$$\sum_{i=1}^{\infty} \operatorname{vol}(I_i) \leqslant m^*\left(\bigcup_{i=1}^{\infty} I_i\right).$$

□

In the one dimensional case every nonempty open set is a union of countably many disjoint open intervals, see [11, Theorem 1.3, p. 6]. By Lemma 1.54 the Lebesgue outer measure of an open set is the sum of volumes of these intervals. Next we consider this question in the higher dimensional case.

A closed dyadic cube is of the form

$$\left[\frac{i_1}{2^k}, \frac{i_1+1}{2^k}\right] \times \cdots \times \left[\frac{i_n}{2^k}, \frac{i_n+1}{2^k}\right], \quad i_1,\dots,i_n \in \mathbb{Z}, \quad k \in \mathbb{Z}.$$

The collection of dyadic cubes $\mathscr{D}_k$, $k \in \mathbb{Z}$, consists of the dyadic cubes with the side length 2^{-k}. The collection of all dyadic cubes in $\mathbb{R}^n$ is

$$\mathscr{D} = \bigcup_{k\in\mathbb{Z}} \mathscr{D}_k.$$

Observe that $\mathscr{D}_k$ consist of cubes whose vertices lie on the lattice $2^{-k}\mathbb{Z}^n$ and whose side length is 2^{-k}. The dyadic cubes in the kth generation can be defined as $\mathscr{D}_k = 2^{-k}([0,1)^n + \mathbb{Z}^n)$. The cubes in $\mathscr{D}_k$ cover the whole $\mathbb{R}^n$ and are pairwise disjoint. Dyadic cubes have a very useful nesting property which states that any two dyadic cubes are either disjoint or one of them is contained in the other.

Lemma 1.55. Every nonempty open set G in $\mathbb{R}^n$ is a union of countably many almost disjoint closed dyadic cubes.

Proof. Consider dyadic cubes in $\mathscr{Q}_1$ that are contained in G and denote

$$\mathscr{Q}_1 = \{Q \in \mathscr{D}_1 : Q \subset G\}.$$

Then consider dyadic cubes in $\mathscr{Q}_2$ that are contained in G and do not intersect any of the cubes in $\mathscr{Q}_1$ and denote

$$\mathscr{Q}_2 = \{Q \in \mathscr{D}_2 : Q \subset G,\ Q \cap J = \emptyset \text{ for every } J \in \mathscr{Q}_1\}.$$

Recursively define

$$\mathscr{Q}_k = \left\{Q \in \mathscr{D}_k : Q \subset G,\ Q \cap J = \emptyset \text{ for every } J \in \bigcup_{i=1}^{k-1} \mathscr{Q}_i\right\}$$

for every $k = 2,3,\dots$. Then $\mathscr{Q} = \bigcup_{k=1}^{\infty} \mathscr{Q}_k$ is a countable collection of almost disjoint closed cubes.

Claim: $G = \bigcup_{Q\in\mathscr{Q}} Q$.

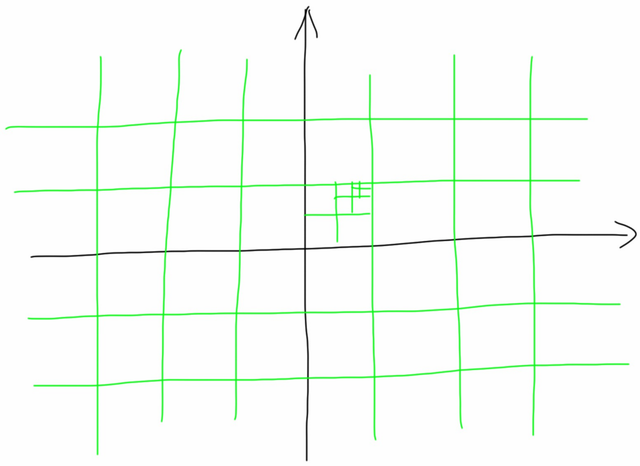

Figure 1.15: Dyadic cubes.

Reason. It is clear from the construction that $\bigcup_{Q\in\mathcal{Q}} Q \subset G$. For the reverse inclusion, let $x \in G$. Let k be so large that the common diameter of the cubes in $\mathcal{D}_k$ is smaller than r, that is, $\sqrt{n}2^{-k} < r$. Since G is open, there exists a ball $B(x,r) \subset G$ with $r > 0$. Since the dyadic $\mathcal{Q}_k$ cubes cover $\mathbb{R}^n$, there exists a dyadic cube $Q \in \mathcal{Q}_k$ with $x \in Q$ and $Q \subset B(x,r) \subset G$. There are two possibilities $Q \in \mathcal{Q}_k$ or $Q \notin \mathcal{Q}_k$. If $Q \in \mathcal{Q}_k$, then $x \in Q \subset \bigcup_{Q\in\mathcal{Q}} Q$. If $Q \notin \mathcal{Q}_k$, there exists $J \in \bigcup_{i=1}^{k-1} \mathcal{Q}_i$ with $J \cap Q \neq \emptyset$. The nesting property of dyadic cubes implies $Q \subset J$ and $x \in Q \subset J \subset \bigcup_{Q\in\mathcal{Q}} Q$. ■

The argument shows that every nonempty open set is a union of countably many disjoint half open dyadic intervals

$$\left[\frac{i_1}{2^k}, \frac{i_1+1}{2^k}\right) \times \cdots \times \left[\frac{i_n}{2^k}, \frac{i_n+1}{2^k}\right), \quad i_1, \dots, i_n \in \mathbb{Z}, \quad k \in \mathbb{Z}.$$

In most of the cases we are not interested in the precise value of the Lebesgue outer measure of a set $A \subset \mathbb{R}^n$. Instead, it is enough to know whether $m^*(A) = 0$, $0 < m^*(A) < \infty$ or $m^*(A) = \infty$.

Remark 1.56. A set $A \subset \mathbb{R}^n$ is of Lebesgue outer measure zero if and only if for every $\varepsilon > 0$ there are exist intervals I_i, $i = 1, 2, \dots$, such that $A \subset \bigcup_{i=1}^{\infty} I_i$ and

$$\sum_{i=1}^{\infty} \text{vol}(I_i) < \varepsilon.$$

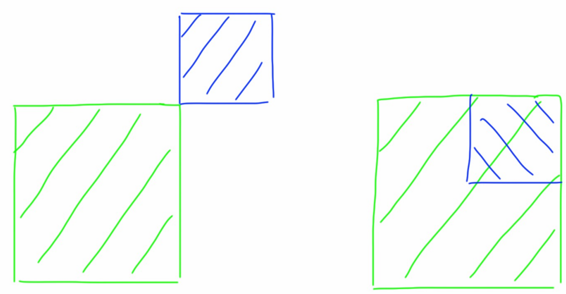

Figure 1.16: The nesting property of the dyadic cubes.

Examples 1.57:

(1) Any one point set is of Lebesgue measure zero, that is, $m^*(\{x\}) = 0$ for every $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. We give two ways to prove the claim.

$\boxed{(1)}$ Let $\varepsilon > 0$ and

$$Q = \left[x_1 - \frac{\varepsilon^{1/n}}{2}, x_1 + \frac{\varepsilon^{1/n}}{2}\right] \times \cdots \times \left[x_n - \frac{\varepsilon^{1/n}}{2}, x_n + \frac{\varepsilon^{1/n}}{2}\right].$$

Observe that Q is a cube with center x and all side lengths equal to $\varepsilon^{1/n}$. A cube is an n-dimensional interval whose side lengths are equal. Then

$$m^*(\{x\}) \leqslant \operatorname{vol}(Q) = \varepsilon,$$

which implies that $m^*(\{x\}) = 0$.

$\boxed{(2)}$ We can cover $\{x\}$ by the degenerate interval $[x_1, x_1] \times \cdots \times [x_n, x_n]$ with zero volume and conclude the claim from this.

(2) Any countable set $A = \{x_1, x_2, \dots\}$, $x_i \in \mathbb{R}^n$ is of Lebesgue measure zero. We give two ways to prove the claim.

$\boxed{(1)}$ Let $\varepsilon > 0$ and Q_i, $i = 1, 2, \dots$, be a closed n-dimensional cube with center x_i and side length $(\frac{\varepsilon}{2^i})^{1/n}$. Then

$$m^*(A) \leqslant \sum_{i=1}^{\infty} \operatorname{vol}(Q_i) \leqslant \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon,$$

which implies that $m^*(A) = 0$.

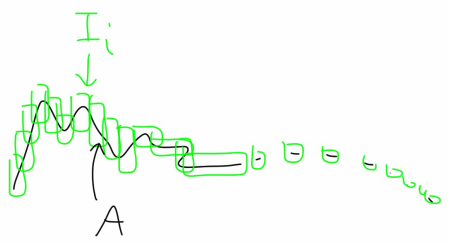

Figure 1.17: A set of Lebesgue measure zero.

$\boxed{(2)}$ By subadditivity

$$m^*(A) = m^*\left(\bigcup_{i=1}^{\infty}\{x_i\}\right) \leqslant \sum_{i=1}^{\infty} m^*(\{x_i\}) = 0.$$

(3) Let $A = \{x = (x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\} \subset \mathbb{R}^2$. Then the 2-dimensional Lebesgue measure of A is zero.

Reason. Let $A_i = \{x = (x_1, x_2) \in \mathbb{R}^2 : i \leqslant x_1 < i+1, x_2 = 0\}$, $i \in \mathbb{Z}$. Then $A = \bigcup_{i \in \mathbb{Z}} A_i$. Let $\varepsilon > 0$ and $I = [i, i+1] \times \left[-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}\right]$. Then $A_i \subset I$ and $\mathrm{vol}(I) = \varepsilon$. This implies $m^*(A_i) = 0$ and

$$m^*(A) \leqslant \sum_{i \in \mathbb{Z}} m^*(A_i) = 0.$$

■

(4) $m^*(\mathbb{R}^n) = \infty$.

Reason. Let I_i, $i = 1, 2, \ldots$, be a collection of closed intervals such that $\mathbb{R}^n \subset \bigcup_{i=1}^{\infty} I_i$. Consider the cubes $Q_j = [-j, j]^n = [-j, j] \times \cdots \times [-j, j]$, $j = 1, 2, \ldots$. Then $Q_j \subset \bigcup_{i=1}^{\infty} I_i$ and by Lemma 1.52 the Lebesgue outer measure of a closed interval coincides with its volume. Thus we have

$$(2j)^n = \mathrm{vol}(Q_j) = m^*(Q_j) \leqslant \sum_{i=1}^{\infty} \mathrm{vol}(I_i).$$

Letting $j \to \infty$, we see that $\sum_{i=1}^{\infty} \mathrm{vol}(I_i) = \infty$ for every covering. This implies that $m^*(\mathbb{R}^n) = \infty$.

■

(5) Every nonempty open set has positive Lebesgue outer measure.

Reason. Let $G \subset \mathbb{R}^n$ be open. Then for every $x \in G$, there exists a ball $B(x,r) \subset G$ with $r > 0$. The ball $B(x,r)$ contains the cube Q with the center x and diameter $\frac{r}{2}$. On the other hand, the $\operatorname{diam}(Q) = \sqrt{n}\, l(Q)$, where $l(Q)$ is the side lenght of Q. From this we conclude that $l(Q) = r/(2\sqrt{n})$ and thus

$$Q = \left[x_1 - \frac{r}{4\sqrt{n}}, x_1 + \frac{r}{4\sqrt{n}}\right] \times \cdots \times \left[x_n - \frac{r}{4\sqrt{n}}, x_n + \frac{r}{4\sqrt{n}}\right].$$

By Lemma 1.52 this implies

$$m^*(G) \geqslant m^*(Q) = \operatorname{vol}(Q) = \left(\frac{r}{2\sqrt{n}}\right)^n > 0.$$

■

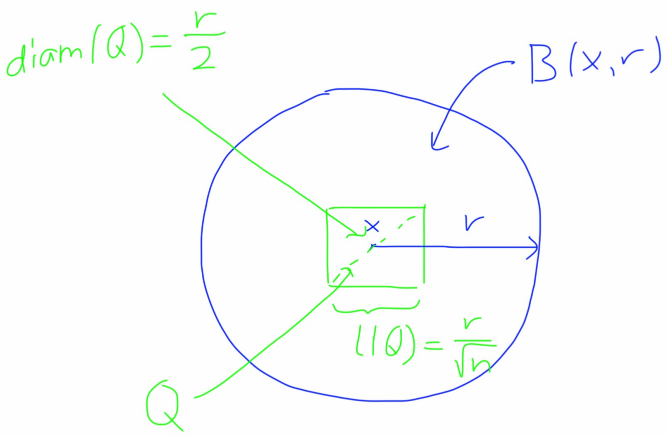

Figure 1.18: A cube inside a ball.

Observe, that every nonempty open set contains uncountably many points, since all countable sets have Lebesgue measure zero.

1.8 Invariance properties of the Lebesgue measure

The following invariance properties of the Lebesgue measure follow from the corresponding properties of the volume of an interval.

(1) (Translation invariance) Let $A \subset \mathbb{R}^n$, $x_0 \in \mathbb{R}^n$ and denote $A + x_0 = \{x + x_0 \in \mathbb{R}^n : x \in A\}$. Then

$$m^*(A + x_0) = m^*(A).$$

This means that the Lebesgue outer measure is invariant in translations.

Reason. Intervals are mapped to intervals in translations and

$$A \subset \bigcup_{i=1}^{\infty} I_i \quad \Longleftrightarrow \quad A + x_0 \subset \bigcup_{i=1}^{\infty} (I_i + x_0).$$

Clearly $\operatorname{vol}(I_i) = \operatorname{vol}(I_i + x_0)$, $i = 1, 2, \dots$, and thus

$$\begin{aligned} m^*(A + x_0) &= \inf\left\{ \sum_{i=1}^{\infty} \operatorname{vol}(I_i + x_0) : A + x_0 \subset \bigcup_{i=1}^{\infty} (I_i + x_0) \right\} \\ &= \inf\left\{ \sum_{i=1}^{\infty} \operatorname{vol}(I_i) : A \subset \bigcup_{i=1}^{\infty} I_i \right\} = m^*(A). \end{aligned}$$

Moreover, A is Lebesgue measurable if and only if $A + x_0$ is Lebesgue measurable. To see this, assume that A is Lebesgue measurable. Then

$$\begin{aligned} &m^*(E \cap (A + x_0)) + m^*(E \setminus (A + x_0)) \\ &= m^*\big(((E - x_0) \cap A) + x_0\big) + m^*\big(((E - x_0) \setminus A) + x_0\big) \\ &= m^*((E - x_0) \cap A) + m^*((E - x_0) \setminus A) \quad \text{(translation invariance)} \\ &= m^*(E - x_0) \quad (A \text{ is measurable}) \\ &= m^*(E) \quad \text{(translation invariance)} \end{aligned}$$

for every $E \subset \mathbb{R}^n$. This shows that $A + x_0$ is Lebesgue measurable. The equivalence follows from this. This claim can also be proved using Theorem 1.44 or Theorem 1.59. ■

(2) (Reflection invariance) Let $A \subset \mathbb{R}^n$ and denote $-A = \{-x \in \mathbb{R}^n : x \in A\}$. Then

$$m^*(-A) = m^*(A).$$

This means that the Lebesgue outer measure is invariant in reflections.

(3) (Scaling property) Let $A \subset \mathbb{R}^n$, $\delta \geqslant 0$ and denote $\delta A = \{\delta x \in \mathbb{R}^n : x \in A\}$. Then

$$m^*(\delta A) = \delta^n m^*(A).$$

This shows that the Lebesgue outer measure behaves as a volume is expected in dilations.

(4) (Change of variables) Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a general linear mapping. Then

$$m^*(L(A)) = |\det L| m^*(A).$$

This is a change of variables formula, see [7] pages 65–80 or [16] pages 612–619. Moreover, if A is Lebesgue measurable, then $L(A)$ is Lebesgue measurable. However, if $L : \mathbb{R}^n \to \mathbb{R}^m$ with $m < n$, then $L(A)$ need not be Lebesgue measurable. We shall return to this question later.

(5) (Rotation invariance) A rotation is a linear mapping $L:\mathbb{R}^n \to \mathbb{R}^n$ with $LL^* = I$, where L^* is the transpose of L and I is the identity mapping. Since $\det L = \det L^*$ it follows that $|\det L| = 1$. The change of variables formula implies that

$$m^*(L(A)) = m^*(A)$$

and thus the Lebesgue outer measure is invariant in rotations. This also shows that the Lebesgue outer measure is invariant in orthogonal linear mappings $L:\mathbb{R}^n \to \mathbb{R}^n$. Recall that L is orthogonal, if $T^{-1} = T^*$. Moreover, the Lebesgue outer measure is invariant under rigid motions $\Phi:\mathbb{R}^n \to \mathbb{R}^n$, $\Phi(x) = x_0 + Lx$, where L is orthogonal.

THE MORAL: The Lebesgue measure is invariant in rigid motions and is consistent with scalings. Later we shall see that the Lebesgue measure is essentially the only measure with these properties.

Example 1.58. Let $B(x,r) = \{y \in \mathbb{R}^n : |y-x| < r\}$ be a ball with the center $x \in \mathbb{R}^n$ and radius $r > 0$. By the translation invariance

$$m(B(x,r)) = m(B(x',r)) \quad \text{for every} \quad x' \in \mathbb{R}^n$$

and by the scaling property

$$m(B(x,ar)) = a^n m(B(x,r)) \quad \text{for every} \quad a > 0.$$

In particular, $m(B(x,r)) = r^n m(B(0,1))$ for every $r > 0$. Thus the Lebesgue measure of any ball is uniquely determined by the measure of the unit ball. This question will be discussed further in Example 3.36.

1.9 Lebesgue measurable sets

Next we discuss measurable sets for the Lebesgue outer measure. We have already shown that the Lebesgue outer measure is a Radon outer measure, see the discussion after Lemma 1.50. In particular, all Borel sets are Lebesgue measurable.

THE MORAL: Open and closed sets are Lebesgue measurable and all sets obtained from these sets by countably many set theoretic operations, as complements intersections and unions, are Lebesgue measurable set. Thus the majority of sets that we actually encounter in real analysis will be Lebesgue measurable. However, there is a set which is not Lebesgue measurable, as we shall see soon.

We revisit approximation properties of Lebesgue measurable sets that were already know from Theorem 1.44 and Corollary 1.46. Certain arguments are easier for the Lebesgue outer measure than for a general Radon outer measure.

Theorem 1.59. If $A \subset \mathbb{R}^n$ is Lebesgue measurable, then the following claims are true.

(1) For every $\varepsilon > 0$, there exists an open $G \supset A$ such that $m^*(G \setminus A) < \varepsilon$.

(2) For every $\varepsilon > 0$, there exists a closed $F \subset A$ such that $m^*(A \setminus F) < \varepsilon$.

(3) If $m^*(A) < \infty$, for every $\varepsilon > 0$, there exists a compact set $K \subset A$ such that $m^*(A \setminus K) < \varepsilon$.

(4) $m^*(A) = \inf\{m^*(G) : A \subset G,\ G \text{ open}\}$. This holds for every $A \subset \mathbb{R}^n$.

(5) $m^*(A) = \sup\{m^*(K) : K \subset A,\ K \text{ compact}\}$.

Proof. $\boxed{(1)}$ Assume that $m^*(A) < \infty$. Let $\varepsilon > 0$. Let I_i, $i = 1, 2, \dots$, be open intervals such that $A \subset \bigcup_{i=1}^\infty I_i$ and

$$\sum_{i=1}^\infty \operatorname{vol}(I_i) < m^*(A) + \varepsilon.$$

Let $G = \bigcup_{i=1}^\infty I_i$. By subadditivity $m^*(G) < m^*(A) + \varepsilon$. Since A is Lebesgue measurable and $A \subset G$, we have

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

This also follows from additivity on pairwise disjoint measurable sets. This implies

$$m^*(G \setminus A) = m^*(G) - m^*(A) < \varepsilon.$$

In the case $m^*(A) = \infty$ we consider the exhaustion $A = \bigcup_{i=1}^\infty (A \cap B(0,i))$ as in the proof of Theorem 1.44.

$\boxed{(2)}$ Let $\varepsilon > 0$. By Lemma 1.9 the set $\mathbb{R}^n \setminus A$ is measurable and by (1) there exists an open set $G \supset R^n \setminus A$ with $m^*(G \setminus (\mathbb{R}^n \setminus A)) < \varepsilon$. Let $F = \mathbb{R}^n \setminus G$. Then F is closed, $F \subset A$, $A \setminus F = G \setminus (\mathbb{R}^n \setminus A)$ and

$$m^*(A \setminus F) = m^*(G \setminus (\mathbb{R}^n \setminus A)) < \varepsilon.$$

$\boxed{(3)}$ Let $\varepsilon > 0$. By (2) there is a closed set $F \subset A$ such that $m^*(A \setminus F) < \varepsilon$. Consider the closed balls centered at the origin $\overline{B}(0,i) = \{x \in \mathbb{R}^n : |x| \leqslant i\}$,and let $K_i = F \cap \overline{B}(0,i)$ for $i = 1, 2, \dots$. The sets K_i, $i = 1, 2, \dots$, are compact as a closed and bounded set. Then $A \setminus K_i$, $i = 1, 2, \dots$, is a decreasing sequence of measurable sets with $\bigcap_{i=1}^\infty (A \setminus K_i) = A \setminus K$. Since $m^*(A) < \infty$, by Theorem 1.20

$$\lim_{i\to\infty} m^*(A \setminus K_i) = m^*\left(\bigcap_{i=1}^\infty (A \setminus K_i)\right) = m^*(A \setminus K) < \varepsilon.$$

This implies that $m^*(A \setminus K_i) < \varepsilon$ for large enough i.

$\boxed{(4)}$ Let $\varepsilon > 0$. Take cubes Q_i, $i = 1, 2, \dots$, such that $A \subset \bigcup_{i=1}^\infty Q_i$ and

$$\sum_{i=1}^\infty \operatorname{vol}(Q_i) < m^*(A) + \frac{\varepsilon}{2}.$$

Let Q_i^0 denote an open cube containing Q_i with

$$\operatorname{vol}(Q_i^0) \leqslant \operatorname{vol}(Q_i) + \frac{\varepsilon}{2^{i+1}}$$

for $i = 1, 2, \dots$. Then $G = \bigcup_{i=1}^{\infty} Q_i^0$ is open and

$$\begin{aligned}
m^*(G) &\leqslant \sum_{i=1}^{\infty} m^*(Q_i^0) = \sum_{i=1}^{\infty} \operatorname{vol}(Q_i^0) \\
&\leqslant \sum_{i=1}^{\infty} \left(\operatorname{vol}(Q_i) + \frac{\varepsilon}{2^i} \right) \leqslant \sum_{i=1}^{\infty} \operatorname{vol}(Q_i) + \frac{\varepsilon}{2} \\
&< m^*(A) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = m^*(A) + \varepsilon.
\end{aligned}$$

$\boxed{(5)}$ The claim follows as in Corollary 1.46. □

By the previous approximation properties we obtain a characterization of Lebesgue measurable sets. Recall that a G_δ set is a countable intersection of open sets and a F_σ set is a countable union of closed sets. In particular, G_δ and F_σ sets are Borel sets. The following characterization of Lebesgue measurable sets is a reformulation of Remark 1.45.

Corollary 1.60. The following claims are equivalent for a set $A \subset \mathbb{R}^n$:

(1) A is Lebesgue measurable,

(2) A is a G_δ set with a set of measure zero removed and

(3) A is a union of a F_σ set and a set of measure zero.

THE MORAL: An arbitrary Lebesgue measurable set differs from a Borel set only by a set of measure zero.

Does there exist Lebesgue measurable sets that are not Borel sets? We shall see in Section 2.3, that

(1) there are Lebesgue measurable sets that are not Borel sets and

(2) the restriction of the Lebesgue measure to the Borel sets is not a complete measure.

Lebesgue measurable sets arise as a completion of the σ-algebra of Borel sets, that is, adding all sets of measure zero as in Remark 1.15.

Remarks 1.61:

(1) There are number of equivalent ways of defining Lebesgue measurability. Probably the simplest definition states that a set $A \subset \mathbb{R}^n$ is Lebesgue measurable, if for every $\varepsilon > 0$ there exists an open set G with $G \supset A$ and

$$m^*(G \setminus A) < \varepsilon.$$

Compare this carefully to Theorem 1.59 (5), which holds for all sets $A \subset \mathbb{R}^n$, see also Remark 1.47.

Reason. First assume that $A \subset \mathbb{R}^n$ is Lebesgue measurable. By Theorem 1.59, for every $\varepsilon > 0$, there exists an open set G with $G \supset A$ and $m^*(G \setminus A) < \varepsilon$.

Assume then that $A \subset \mathbb{R}^n$ is such that $\varepsilon > 0$ there exists an open set G with $G \supset A$ and $m^*(G \setminus A) < \varepsilon$. In particular, for every $i = 1, 2, \dots$, there exists an open set $G_i \supset A$ with $m^*(G_i \setminus A) < \frac{1}{i}$. Then $B = \bigcap_{i=1}^{\infty} G_i$ is a G_δ set that contains A and $B \setminus A \subset G_i \setminus A$ for every $i = 1, 2, \dots$. This implies

$$m^*(B \setminus A) \leqslant m^*(G_i \setminus A) < \frac{1}{i}$$

for every $i = 1, 2, \dots$ and thus $m^*(B \setminus A) = 0$. The set $B \setminus A$ is Lebesgue measurable as a set of measure zero and thus $A = B \setminus (B \setminus A)$ is Lebesgue measurable as a union of two Lebesgue measurable sets. ■

(2) Alternatively a set $A \subset \mathbb{R}^n$ is Lebesgue measurable, if for every $\varepsilon > 0$ there exists an closed set F with $F \supset A$ and $m^*(A \setminus F) < \varepsilon$. Observe that the same argument applies to a general Radon outer measure, see Remark 1.45.

Remark 1.62. The Lebesgue outer measure has a product structure. Let $n, m \geqslant 1$ be natural numbers.

(1) If $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$, then $(m^{n+m})^*(A \times B) \leqslant (m^n)^*(A)(m^m)^*(B)$.
(2) If $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ are Lebesgue measurable sets, then $A \times B$ is Lebesgue measurable and $(m^{n+m})^*(A \times B) = (m^n)^*(A)(m^m)^*(B)$.

We return to this later.

Definition 1.63. The Lebesgue measure is defined to be the Lebesgue outer measure on the σ-algebra of Lebesgue measurable sets. We denote the Lebesgue measure by m. In particular, the Lebesgue measure is countably additive on pairwise disjoint Lebesgue measurable sets.

Remark 1.64. The Lebesgue measure is the unique in the sense that it is the only mapping $A \mapsto \mu(A)$ from the Lebesgue measurable sets to $[0, \infty]$ satisfying the following conditions.

(1) $\mu(\emptyset) = 0$.
(2) (Countable additivity) If $A_i \subset \mathbb{R}^n$, $i = 1, 2, \dots$, are pairwise disjoint Lebesgue measurable sets, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.
(3) (Translation invariance) If A is a Lebesgue measurable set and $x \in \mathbb{R}^n$, then $\mu(A + x) = \mu(A)$.
(4) (Normalisation) $\mu((0, 1))^n) = 1$.

Reason. Subdivide the open unit cube $Q = (0,1)^n$ to a union of 2^{kn} pairwise disjoint half open dyadic intervals Q_i of side length 2^{-k}, $k = 1,2,\dots$. By translation invariance all cubes Q_i have the same measure, that is, $m(Q_i) = m(Q_j)$ for $i, j = 1,2\dots$, and by countable additivity on disjoint measurable sets the sum of their measures equals the measure of the entire cube Q which, by normalisation, has measure 1. This implies

$$2^{kn}m(Q_i) = \sum_{i=1}^{2^{kn}} m(Q_i) = m(Q) = 1 = \mu(Q) = \sum_{i=1}^{2^{kn}} \mu(Q_i) = 2^{kn}\mu(Q_i).$$

A similar argument can be done for $k = 0,-1,-2,\dots$ and thus $\mu(Q) = m(Q)$ for all dyadic cubes $Q \subset \mathbb{R}^n$. Since every open set can be represented as a union of pairwise disjoint half open dyadic cubes, additivity implies $\mu(G) = m(G)$ for all open sets $G \subset \mathbb{R}^n$. ∎

1.10 A nonmeasurable set

The Lebesgue outer measure m^* on $\mathbb{R}^n$ measuring the n-dimensional volume of subsets of $\mathbb{R}^n$ would ideally have the following properties:

(1) $m^*(A)$ is defined for every set $A \subset \mathbb{R}^n$,

(2) m^* is an outer measure,

(3) m^* is countably additive on pairwise disjoint sets: $m^*(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m^*(A_i)$ for pairwise disjoint sets $A_i \subset \mathbb{R}^n$, $i = 1,2,\dots$.

(4) m^* is translation invariant: $m^*(A) = m^*(A+x)$ for $x \in \mathbb{R}$

However, it is impossible to satisfy all these are simultaneously if we assume the axiom of choice. The Lebesgue outer measure satisfies (1), (2) and (4). The Lebesgue outer measure also satisfies (3), if all sets A_i., $i = 1,2,\dots$, are Lebesgue measurable, but additivity may break down for nonmeasurable sets. We have shown that Borel sets are Lebesgue measurable, but we have not yet ruled out the possibility that every set is Lebesgue measurable. Next we shall show that there exists a nonmeasurable set for the Lebesgue outer measure on $\mathbb{R}$. Such a set can be constructed using the axiom of choice. For the role of the axiom of choice in the construction of a nonmeasurable set, we refer to [6, Chapter 33].

Remark 1.65. The axiom of choice states that if E is a set and $\{E_\alpha\}$ is a collection of nonempty subsets of E, then there exists a function (a choice function) $\alpha \mapsto x_\alpha$ such that $x_\alpha \in E_\alpha$ for every α. The indexing set of α's is not assumed to be countable.

THE MORAL: The axiom of choice states that we have a set which contains exactly one point from each set in an uncountable collection of sets.

Theorem 1.66. There exists a set $E \subset [0,1]$ which is not Lebesgue measurable.

STRATEGY: We show that there exists a set $E \subset [0,1]$ of positive Lebesgue outer measure such that the translated sets $E+q$, $q \in [-1,1]\cap\mathbb{Q}$, form a pairwise disjoint covering of the interval $[0,1]$. As the Lebesgue measure is translation invariant and countably additive on measurable sets, the set E cannot be Lebesgue measurable.

Proof. Define an equivalence relation on the real line by

$$x \sim y \Leftrightarrow x - y \in \mathbb{Q}.$$

Claim: $\sim$ is an equivalence relation.

Reason. It is clear that $x \sim x$ and that if $x \sim y$ then $y \sim x$. To prove transitivity, assume that $x \sim y$ and $y \sim z$. Then $x - y = q_1$ and $y - z = q_2$, where $q_1,\ q_2 \in \mathbb{Q}$ and

$$x - z = (x-y) + (y-z) = q_1 + q_2 \in \mathbb{Q}.$$

This implies that $x \sim z$. ■

The equivalence relation $\sim$ decomposes $\mathbb{R}$ into disjoint equivalence classes. Denote the equivalence class containing x by E_x. Note that if $x \in \mathbb{Q}$, then $E_x = \mathbb{Q}$. Note also that each equivalence class is countable and therefore, since $\mathbb{R}$ is uncountable, there must be an uncountable number of equivalence classes. Each equivalence class is dense in $\mathbb{R}$ and has a nonempty intersection with $[0,1]$. By the axiom of choice, there is a set E which consist of precisely one element of each equivalence class belonging to $[0,1]$. If x and y are arbitrary elements of E, then $x - y$ is an irrational number, for otherwise they would belong to the same equivalence class, contrary to the definition of E.

We claim that E is not Lebesgue measurable. Assume for a contradiction that E is Lebesgue measurable. Then the translated sets

$$E + q = \{x + q : x \in E\}, \quad q \in \mathbb{Q},$$

are Lebesgue measurable.

Claim: The sets $E+q$ are disjoint, that is,

$$(E+q)\cap(E+r) = \emptyset \quad \text{whenever} \quad q,r \in \mathbb{Q}, q \neq r.$$

Reason. For the contradiction, assume that $y \in (E+q)\cap(E+r)$ with $q \neq r$. Then $y = x + q$ and $y = z + r$ for some $x,z \in E$. Thus

$$x - z = (y-q) - (y-r) = r - q \in \mathbb{Q},$$

which implies that $x \sim z$. Since E contains exactly one element of each equivalence class, we have $x = z$ and consequently $r = q$. ■

Claim: $[0,1] \subset \bigcup_{q\in[-1,1]\cap\mathbb{Q}} (E+q) \subset [-1,2]$.

Reason. Let $x \in [0,1]$ and let y be the representative of the equivalence class E_x belonging to E. In particular, $x \sim y$ from which it follows that $x - y \in \mathbb{Q}$. Denote $q = x - y$. Since $x, y \in [0,1]$ we have $q \in [-1,1]$ and $x = y + q \in E + q$. This proves the first inclusion. The second inclusion is clear. ■

Since the translated sets $E+q$, $q \in [-1,1]\cap\mathbb{Q}$, are pairwise disjoint and Lebesgue measurable, by countable additivity and translation invariance,

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

which is 0 if $m^*(E) = 0$ and ∞ if $m^*(E) > 0$. On the other hand, since $[0,1] \subset \bigcup_{q\in[-1,1]\cap\mathbb{Q}}(E+q)$, and monotonicity

$$\sum_{q\in[-1,1]\cap\mathbb{Q}} m^*(E) = m^*\left(\bigcup_{q\in[-1,1]\cap\mathbb{Q}} (E+q)\right) \geqslant m^*([0,1]) = 1 > 0,$$

which implies $m^*(E) > 0$ and, consequently,

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

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

$$\infty = m^*\left(\bigcup_{q\in[-1,1]\cap\mathbb{Q}} (E+q)\right) \leqslant m^*([-1,2]) = 3.$$

This is a contradiction and thus E cannot be Lebesgue measurable. □

Remark 1.67. The proof shows that the $E \subset [0,1]$ is not Lebesgue measurable, $m^*(E) > 0$, the sets $E+q$, $q \in [-1,1]\cap\mathbb{Q}$, are pairwise disjoint,

$$m^*\left(\bigcup_{q\in[-1,1]\cap\mathbb{Q}} (E+q)\right) \leqslant m^*([-1,2]) = 3 < \infty$$

and

$$\sum_{q\in[-1,1]\cap\mathbb{Q}} m^*(E+q) = \sum_{q\in[-1,1]\cap\mathbb{Q}} m^*(E) = \infty.$$

Thus

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

and countable additivity on pairwise disjoint sets fails.

Remark 1.68. By a modification of the above proof we see that any set $A \subset \mathbb{R}$ with $m^*(A) > 0$ contains a set B which is not Lebesgue measurable.

Reason. Let $A \subset \mathbb{R}$ be a set with $m^*(A) > 0$. Then there must be at least one interval $[i, i+1]$, $i \in \mathbb{Z}$, such that $m^*(A \cap [i, i+1]) > 0$, otherwise

$$m^*(A) = m^*\left(\bigcup_{i\in\mathbb{Z}}(A \cap [i, i+1])\right) \leqslant \sum_{i\in\mathbb{Z}} m^*(A \cap [i, i+1]) = 0.$$

By a translation, we may assume that $m^*(A \cap [0,1]) > 0$ and $A \subset [0,1]$. By the notation of the proof of the previous theorem,

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

Again, by countable subadditivity, at least one of the sets $(E+q) \cap A$, $q \in [-1,1]$, has positive Lebesgue outer measure. Set $B = (E+q) \cap A$ with $m^*(B) > 0$.

The same argument as in the proof of the previous theorem shows that B is not Lebesgue measurable. Indeed, assume that B is Lebesgue measurable. Since the translated sets $B+q$, $q \in [-1,1] \cap \mathbb{Q}$, are disjoint and Lebesgue measurable, by countable additivity and translation invariance,

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

since $m^*(B) > 0$. On the other hand,

$$m^*\left(\bigcup_{q\in[-1,1]\cap\mathbb{Q}} (B+q)\right) \leqslant m^*([-1,2]) = 3 < \infty.$$

This is a contradiction and thus B cannot be Lebesgue measurable. ■

Remarks 1.69:

(1) A Lebesgue nonmeasurable set is not a Borel set, since all Borel sets are Lebesgue measurable.

(2) The construction above can be used to show that there is a Lebesgue measurable set which is not a Borel set. Let $A \subset \mathbb{R}$ be a set which is not measurable with respect to the one-dimensional Lebesgue measure. Consider

$$B = \{(x, 0, \dots, 0) : x \in A\} \subset \mathbb{R}^n.$$

Then the n-dimensional Lebesgue measure of B is zero and thus B is Lebesgue measurable with respect to the n-dimensional Lebesgue measure. However, B is not a Borel set in $\mathbb{R}^n$, since the projection $\pi : \mathbb{R}^n \to \mathbb{R}$, $\pi(x) = x_1$ maps Borel set to Borel sets. This implies that π does not map measurable sets with respect to the n-dimensional Lebesgue measure to Lebesgue measurable sets with respect to one-dimensional Lebesgue measure. We shall see more examples later.

Remark 1.70. The Banach-Tarski paradox shows that the unit ball in $\mathbb{R}^3$ can be disassembled into a finite number of disjoint pieces (in fact, just five pieces suffice), which can then be reassembled (after translating and rotating each of the pieces) to form two disjoint copies of the original ball. The pieces used in this decomposition are highly pathological and their construction applies the axiom of choice. In particular, the pieces cannot be Lebesgue measurable, because otherwise the additivity fails. Banach-Tarski paradox does not hold in $\mathbb{R}^2$. This is because $\mathbb{R}^2$ has less symmetries compared to $\mathbb{R}^3$ (this has nothing to do with the existence of non-measurable sets, only with the existence of appropriate non-measurable sets).

1.11 The Cantor set

Cantor sets constructed in this section give several examples of unexpected features in analysis. The middle thirds Cantor set is a subset of the interval $C_0 = [0,1]$. The construction will proceed in steps. At the first step, let $I_{1,1}$ denote the open interval $(\frac{1}{3}, \frac{2}{3})$. Then $I_{1,1}$ is the open middle third of C_0. At the second step we denote two open intervals $I_{2,1}$ and $I_{2,2}$ each being the open middle third of one of the two intervals comprising $I \setminus I_{1,1}$ and so forth. At the kth step, we obtain 2^{k-1} pairwise disjoint open intervals $I_{k,i}$, $i = 1, \dots, 2^{k-1}$, and denote

$$C_0 = [0,1], \quad C_k = C_{k-1} \setminus \bigcup_{i=1}^{2^{k-1}} I_{k,i}, \quad k = 1, 2, \dots.$$

Note:

$$C_k = \bigcup_{a_1,\dots,a_k \in \{0,2\}} \left[\sum_{i=1}^{k} \frac{a_i}{3^i}, \sum_{i=1}^{k} \frac{a_i}{3^i} + \frac{1}{3^k} \right].$$

Thus C_k consists of 2^k closed intervals of length $\frac{1}{3^k}$. Let us denote these intervals by $J_{k,i}$, $i = 1, 2, \dots, 2^k$.

The (middle thirds) Cantor set is the intersection of all sets C_k, that is,

$$C = \bigcap_{k=0}^{\infty} C_k.$$

Note:

$$C = \left\{ \sum_{i=1}^{\infty} \frac{a_i}{3^i} : a_i \in \{0,2\}, i = 1, 2, \dots \right\}.$$

Note that C contains more points that the end points

$$\left\{ \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{2}{9}, \frac{7}{9}, \frac{8}{9}, \frac{1}{27}, \dots \right\}$$

of the extracted open intervals. For example, $\frac{1}{4} \in C$, but it is not an end point of any of the intervals (exercise).

Since every C_k, $k = 0, 1, 2, \dots$, is closed, the intersection C is closed. Since C is also bounded, it is a compact subset of $[0,1]$.

Claim: C is uncountable.

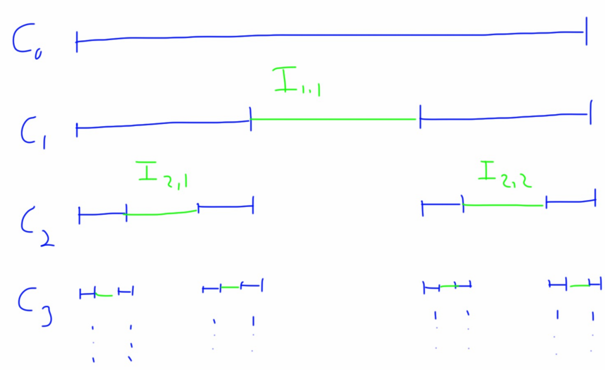

Figure 1.19: The Cantor construction.

Reason. For a contradiction, assume that $C = \{x_1, x_2, \dots\}$ is countable. Let J_1 be one of the closed intervals $J_{1,i}$, $i = 1, 2$, in the first step of construction of the Cantor set with $x_1 \notin J_1$. We continue recursively. Let J_2 be one of the closed intervals $J_{2,i}$, $i = 1, 2, 3, 4$, in the second step of construction of the Cantor set with $x_2 \notin J_2$ and $J_2 \subset J_1$. By continuing this way, we obtain a decreasing sequence of closed intervals $J_{k+1} \subset J_k$, $i = 1, 2, \dots$, such that $C \cap \bigcap_{k=1}^{\infty} J_k = \emptyset$. On the other hand, $\bigcap_{k=1}^{\infty} J_k \neq \emptyset$ and thus there exists a point $x \in \bigcap_{k=1}^{\infty} J_k$. By the definition of the Cantor set, we have $x \in C$, which implies $C \cap \bigcap_{k=1}^{\infty} J_k \neq \emptyset$. This is a contradiction. ■

Moreover, C is nowhere dense and perfect. A set is called nowhere dense if its closure does not have interior points and perfect if it does not have isolated points, that is, every point of the set is a limit point of the set. We show that C is an uncountable set of measure zero.

Claim: $m^*(C) = 0$.

Reason. Since

$$m^*(C_k) = \sum_{i=1}^{2^k} m^*(J_{k,i}) = \sum_{i=1}^{2^k} \left(\frac{1}{3}\right)^k = 2^k \left(\frac{1}{3}\right)^k = \left(\frac{2}{3}\right)^k,$$

by Theorem 1.20 we have

$$m^*(C) = m^*\left(\bigcap_{k=0}^{\infty} C_k\right) = \lim_{k\to\infty} m^*(C_k) = 0.$$

This can be also seen directly from the definition of the Lebesgue measure, since C_k consists of finitely many intervals whose lengths sum up to $\left(\frac{2}{3}\right)^k$. This is arbitrarily small by choosing k large enough. ■

Remark 1.71. Every real number can be represented as a decimal expansion. Instead of using base 10, we may take for example 3 as the base. In particular, every $x \in [0,1]$ can be written as a ternary expansion

$$x = \sum_{i=1}^{\infty} \frac{\alpha_i}{3^i},$$

where $\alpha_i = 0$, 1 or 2 for every $i = 1,2,\dots$ We denote this as $x = .\alpha_1\alpha_2\dots$. In general, this decomposition is not unique. For example,

$$\frac{1}{3} = \sum_{i=2}^{\infty} \frac{2}{3^i}$$

and $\frac{4}{9} = .11000\dots = .10222\dots$. The reason for this is that

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

The ternary expansion is unique except for a certain type of ambiguity. A number has two different expansions if and only if it has a terminating ternary expansion, that is, only finitely many α_i's are nonzero. For example Let us look at the construction of the Cantor set again. At the first stage we remove the middle third $I_{1,1}$. If $\frac{1}{3} < x < \frac{2}{3}$, then $x = .1\alpha_2\alpha_3\dots$. If $x \in [0,1]\setminus I_{1,1}$, then $x = .0\alpha_2\alpha_3\dots$ or $x = .2\alpha_2\alpha_3\dots$ In either case the value of α_1 determines which of the three subintervals contains x. Repeating this argument show that $x \in [0,1]$ belongs to the Cantor middle thirds set if and only if it has a ternary expansion consisting only on 0's and 2's.

The construction of a Cantor type set C can be modified so that at the kth, stage of the construction we remove 2^{k-1} centrally situated open intervals each of length l_k, $k = 1,2,\dots$, with $l_1 + 2l_2 + \dots 2^{k-1}l_k < 1$. If l_k, $k = 1,2,\dots$, are chosen small enough, then

$$\sum_{k=1}^{\infty} 2^{k-1} l_k < 1.$$

In this case, we have

$$0 < 1 - \sum_{k=1}^{\infty} 2^{k-1} l_k = m^*(C) < 1.$$

and C is called a fat Cantor set. Note that C is a compact nowhere dense and perfect set of positive Lebesgue measure. Observe that $U = [0,1]\setminus C$ is an open set with $\partial U = C$ and $m^*(\partial U) = m^*(C) > 0$.

THE MORAL: The boundary of an open set may have positive Lebesgue measure.

See [7, p. 83–86] and [16, p. 85–87] for more on the Cantor set.

The class of measurable functions will play a central role in the integration theory. This class is closed under usual operations and limits, but certain unexpected features occur. Measurable functions can be approximated by simple functions. Egoroff's theorem states that pointwise convergence of a sequence of measurable functions is almost uniform and Lusin's theorem states that a measurable function is almost continuous.

2 Measurable functions

2.1 Calculus with infinities

Throughout the measure and integration theory we encounter $\pm\infty$. One reason for this is that we want to consider sets of infinite measure as $\mathbb{R}^n$ with respect to the Lebesgue measure. Another reason is that we want to consider functions with singularities as $f : \mathbb{R}^n \to [0,\infty]$, $f(x) = |x|^{-\alpha}$ with $\alpha > 0$. Here we use the interpretation that $f(0) = \infty$. In addition, even if we only consider real valued functions, the limes superiors of sequences and sums may be infinite at some points.

We consider the set of extended real numbers $[-\infty,\infty] = \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$. For simplicity, we write ∞ for $+\infty$. We shall use the following conventions for arithmetic operations on $[-\infty,\infty]$. For $a \in \mathbb{R}$, we define

$$a + (\pm\infty) = (\pm)\infty + a = \pm\infty$$

and $(\pm\infty)+(\pm\infty) = \pm\infty$. Subtraction is defined in a similar manner, but $(\pm\infty)+(\mp\infty)$ and $(\pm\infty)-(\pm\infty)$ are undefined. For multiplication, we define

$$a \cdot (\pm\infty) = (\pm\infty)\cdot a = \begin{cases} \pm\infty, & a > 0, \\ 0, & a = 0, \\ \mp\infty, & a < 0, \end{cases}$$

and $(\pm\infty)\cdot(\pm\infty) = +\infty$ and $(\pm\infty)\cdot(\mp\infty) = -\infty$. The operations $\infty\cdot(-\infty)$, $(-\infty)\cdot\infty$ and $(-\infty)\cdot(-\infty)$ are undefined. With these definitions the standard commutative, associative and distributive rules hold in $[-\infty,\infty]$ in the usual manner.

Cancellation properties have to be considered with some care. For example, $a + b = a + c$ implies $b = c$ only when $|a| < \infty$ and $ab = ac$ implies $b = c$ only when $0 < |a| < \infty$. A general fact is that the cancellation is safe if all terms are finite and nonzero in the case of division. Finally we note that with this interpretation, for example, all sums of nonnegative terms $x_i \in [0,\infty]$, $i = 1,2,\dots$, are convergent with

$$\sum_{i=1}^{\infty} x_i \in [0,\infty].$$

We shall use this interpretation without further notice.

2.2 Measurable functions

AGREEMENT: From now on, we shall not distinguish outer measures from measures with the interpretation that an outer measure restricted to measurable sets is a measure.

Consider a function $f : X \to [-\infty,\infty]$. Recall that the preimage of a set $A \subset [-\infty,\infty]$ is

$$f^{-1}(A) = \{x \in X : f(x) \in A\}.$$

The preimage has the properties

$$f^{-1}\left(\bigcap_{i=1}^{\infty} A_i\right) = \bigcap_{i=1}^{\infty} f^{-1}(A_i),\ f^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(A_i)$$

and $f^{-1}(X \setminus A) = X \setminus f^{-1}(A)$ for $A, A_i \subset [-\infty,\infty]$ for $i = 1,2,\dots$.

We begin with a definition of a measurable function .

Definition 2.1. Let μ be a measure on X. The function $f : X \to [-\infty,\infty]$ is μ-measurable, if the set

$$f^{-1}((a,\infty]) = \{x \in X : f(x) > a\}$$

is μ-measurable for every $a \in \mathbb{R}$.

THE MORAL: As we shall see, in the definition of the integral of a function, it is important that all distribution sets are measurable.

Remarks 2.2:

(1) Every continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is Lebesgue measurable.

Reason. Since f is continuous, the set $\{x \in \mathbb{R}^n : f(x) > a\}$ is open for every $a \in \mathbb{R}$. Since the Lebesgue measure is a Borel measure, the set $\{x \in \mathbb{R}^n : f(x) > a\}$ is Lebesgue measurable for every $a \in \mathbb{R}$. ■

(2) The set $A \subset \mathbb{R}^n$ is Lebesgue measurable set if and only if the characteristic function

$$f : \mathbb{R}^n \to \mathbb{R},\ f(x) = \chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \in \mathbb{R}^n \setminus A, \end{cases}$$

is a Lebesgue measurable function.

Reason.

$$\{x \in \mathbb{R}^n : f(x) > a\} = \begin{cases} \mathbb{R}^n, & a < 0, \\ A, & 0 \leqslant a < 1, \\ \emptyset, & a \geqslant 1. \end{cases}$$

■

By considering a set A which is not Lebesgue measurable, see Section 1.10, we conclude that χ_A a nonmeasurable function with respect to the Lebesgue measure.

(3) The previous remark holds for all outer measures. Moreover, a linear combination of finitely many characteristic functions of measurable sets is a measurable function. Such a function is called a simple function, see Defintion 2.31.

(4) If μ is an outer measure for which all sets are μ-measurable, then all functions are μ-measurable. See Remark 1.5 (6).

(5) If the only measurable sets are $\emptyset$ ja X, then only constant functions are measurable. See Remark 1.5 (5).

Remark 2.3. For μ-measurable subset $A \subset X$ and a function $f : A \to [-\infty, \infty]$, we consider the zero extension $\widetilde{f} : X \to [-\infty, \infty]$,

$$\widetilde{f}(x) = \begin{cases} f(x), & x \in A, \\ 0, & x \in X \setminus A. \end{cases}$$

Then f is μ-measurable if and only if $\widetilde{f}$ is μ-measurable.

THE MORAL: A function defined on a subset is measurable if and only if its zero extension to the entire space is measurable. This allows us to consider functions defined on subsets.

Lemma 2.4. Let μ be a measure on X and $f : X \to [-\infty, \infty]$. Then the following claims are equivalent:

(1) f is μ-measurable,

(2) $\{x \in X : f(x) \geqslant a\}$ is μ-measurable for every $a \in \mathbb{R}$,

(3) $\{x \in X : f(x) < a\}$ is μ-measurable for every $a \in \mathbb{R}$,

(4) $\{x \in X : f(x) \leqslant a\}$ is μ-measurable for every $a \in \mathbb{R}$.

Proof. The equivalence follows from the fact that the collection of μ-measurable sets is a σ-algebra, see Lemma 1.9.

$\boxed{(1) \Rightarrow (2)}$ $\{x \in X : f(x) \geqslant a\} = \bigcap_{i=1}^{\infty} \{x \in X : f(x) > a - \frac{1}{i}\}$.

$\boxed{(2) \Rightarrow (3)}$ $\{x \in X : f(x) < a\} = X \setminus \{x \in X : f(x) \geqslant a\}$.

$\boxed{(3) \Rightarrow (4)}$ $\{x \in X : f(x) \leqslant a\} = \bigcap_{i=1}^{\infty} \{x \in X : f(x) < a + \frac{1}{i}\}$.

$\boxed{(4) \Rightarrow (1)}$ $\{x \in X : f(x) > a\} = X \setminus \{x \in X : f(x) \leqslant a\}$. □

Lemma 2.5. A function $f : X \to [-\infty,\infty]$ is μ-measurable if and only if $f^{-1}(\{-\infty\})$ and $f^{-1}(\{\infty\})$ are μ-measurable and $f^{-1}(B)$ is μ-measurable for every Borel set $B \subset \mathbb{R}$.

Remark 2.6. The proof will show that we could require that $f^{-1}(B)$ is μ-measurable for every open set B. This in analogous to the fact that a function is continuous if and only if $f^{-1}(B)$ is open for every open set B.

Proof. $\boxed{\Longrightarrow}$ Note that

$$f^{-1}(\{-\infty\}) = \{x \in X : f(x) = -\infty\} = \bigcap_{i=1}^{\infty}\{x \in X : f(x) < -i\}$$

and

$$f^{-1}(\{\infty\}) = \{x \in X : f(x) = \infty\} = \bigcap_{i=1}^{\infty}\{x \in X : f(x) > i\}$$

are μ-measurable.

Let

$$\mathscr{F} = \{B \subset \mathbb{R} : B \text{ is a Borel set and } f^{-1}(B) \text{ is } \mu\text{-measurable}\}$$

Claim: $\mathscr{F}$ is a σ-algebra.

Reason. Clearly $\emptyset \in \mathscr{F}$. If $B \in \mathscr{F}$, then $f^{-1}(\mathbb{R} \setminus B) = X \setminus f^{-1}(B)$ is μ-measurable and thus $\mathbb{R} \setminus B \in \mathscr{F}$. If $B_i \in \mathscr{F}$, $i = 1,2,\dots$, then

$$f^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} f^{-1}(B_i)$$

is μ-measurable and thus $\bigcup_{i=1}^{\infty} B_i \in \mathscr{F}$. ■

Then we show that $\mathscr{F}$ contains all open subsets of $\mathbb{R}$. Since every open set in $\mathbb{R}$ is a countable union of pairwise disjoint open intervals and $\mathscr{F}$ is a σ-algebra, it is enough to show that every open interval $(a,b) \in \mathscr{F}$. Now

$$f^{-1}((a,b)) = f^{-1}([-\infty,b)) \cap f^{-1}((a,\infty])$$

where $f^{-1}([-\infty,b)) = \{x \in X : f(x) < b\}$ and $f^{-1}((a,\infty]) = \{x \in X : f(x) > a\}$ are μ-measurable. This implies that $f^{-1}((a,b))$ is μ-measurable. Since $\mathscr{F}$ is a σ-algebra that contains open sets, it also contains the Borel sets.

$\boxed{\Longleftarrow}$ Let $B = (a,\infty)$ with $a \in \mathbb{R}$. Then

$$\{x \in X : f(x) > a\} = f^{-1}(B \cup \{\infty\}) = f^{-1}(B) \cup f^{-1}(\{\infty\})$$

is μ-measurable. □

As we shall see in Section 2.3, a composed function of two measurable functions is not measurable, in general.

Lemma 2.7. Let μ be a measure on X. If $f : X \to \mathbb{R}$ is μ-measurable and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then the composed function $g \circ f$ is measurable.

Proof. By Lemma 2.5 and Remark 2.6, it is enough to show that the preimage $(g \circ f)^{-1}(B)$ of every open set $B \subset \mathbb{R}$ is μ-measurable. Note that $(g \circ f)^{-1}(B) = f^{-1}(g^{-1}(B))$, since

$$\begin{aligned} x \in (g \circ f)^{-1}(B) &\iff (g \circ f)(x) \in B \iff g(f(x)) \in B \\ &\iff f(x) \in g^{-1}(B) \iff x \in f^{-1}(g^{-1}(B)). \end{aligned}$$

Since g is continuous, the preimage $g^{-1}(B)$ of an open set B is open. Since f is μ-measurable, the preimage $f^{-1}(g^{-1}(B))$ of an open set $g^{-1}(B)$ is μ-measurable. Thus $(g \circ f)^{-1}(B)$ is a μ-measurable set and $g \circ f$ is a μ-measurable function □

Remark 2.8. In fact, it is enough to assume in Lemma 2.7 that g is a Borel function, that is, the preimage of every Borel set is a Borel set.

Reason. Let $A \subset \mathbb{R}$ be a Borel set and $g : X \to \mathbb{R}$ be a Borel function. Then $g^{-1}(A)$ is a Borel set. Since f is μ-measurable, Lemma 2.5 implies that the preimage $f^{-1}(g^{-1}(B))$ is μ-measurable. ■

Remark 2.9. We briefly discuss an abstract version of a definition of a measurable function.

(1) Assume that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are measure spaces and $f : X \to Y$ is a function. Then f is said to be measurable with respect to σ-algebras $\mathcal{M}$ and $\mathcal{N}$ if $f^{-1}(A) \in \mathcal{M}$ whenever $A \in \mathcal{N}$. In our approach we consider $Y = [-\infty, \infty]$ and $\mathcal{N}$ equals the Borel sets in $[-\infty, \infty]$ (endowed with the order topology, see [15]).

(2) Let $(X, \mathcal{M}, \mu)$, $(Y, \mathcal{N}, \nu)$ and $(Z, \mathcal{P}, \gamma)$ be abstract measure spaces. If $f : X \to Y$ and $g : Y \to Z$ are measurable functions in the sense of (1), then the composed function $g \circ f$ is measurable. This follows directly from the abstract definition of measurablity. We might be tempted to conclude that the composed function of Lebesgue measurable functions is Lebesgue measurable. This is not always the case, since the preimage of a Lebesgue measurable set is not necessarily Lebesgue measurable, see Section 2.3. This means that we cannot replace Borel sets by measurable sets in Lemma 2.5.

Lemma 2.10. If $f, g : X \to [-\infty, \infty]$ are μ-measurable functions, then

$$\{x \in X : f(x) > g(x)\}$$

is a μ-measurable set.

Proof. Since the set of rational numbers is countable , we have $\mathbb{Q} = \bigcup_{i=1}^{\infty}\{q_i\}$. If $f(x) > g(x)$, there exists $q_i \in \mathbb{Q}$ such that $f(x) > q_i > g(x)$. This implies that

$$\{x \in X : f(x) > g(x)\} = \bigcup_{i=1}^{\infty} (\{x \in X : f(x) > q_i\} \cap \{x \in X : g(x) < q_i\})$$

is a μ-measurable set. □

Remark 2.11. The sets

$$\{x \in X : f(x) \leqslant g(x)\} = X \setminus \{x \in X : f(x) > g(x)\}$$

and

$$\{x \in X : f(x) = g(x)\} = \{x \in X : f(x) \leqslant g(x)\} \cap \{x \in X : f(x) \geqslant g(x)\}$$

are μ-measurable as well.

Let $f : X \to [-\infty,\infty]$. The positive part of f is

$$f^+(x) = \max(f(x),0) = f(x)\chi_{\{x\in X : f(x)\geqslant 0\}} = \begin{cases} f(x), & f(x) \geqslant 0, \\ 0, & f(x) < 0, \end{cases}$$

and the negative part is

$$f^-(x) = -\min(f(x),0) = -f(x)\chi_{\{x\in X : f(x)\leqslant 0\}} = \begin{cases} -f(x), & f(x) \leqslant 0, \\ 0, & f(x) > 0. \end{cases}$$

Observe that $f^+, f^- \geqslant 0$, $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Splitting a function into positive and negative parts will be a useful tool in measure theory.

Lemma 2.12. A function $f : X \to [-\infty,\infty]$ is μ-measurable if and only if f^+ and f^- are μ-measurable.

Proof. $\boxed{\Longrightarrow}$ Assume that f is a μ-measurable function. Then

$$\{x \in X : f^+(x) > a\} = \begin{cases} \{x \in X : f(x) > a\}, & a \geqslant 0, \\ X, & a < 0, \end{cases}$$

is a μ-measurable set. This implies that f^+ is a μ-measurable function. Moreover, $f^- = (-f)^+$.

$\boxed{\Longleftarrow}$ Assume that f^+ and f^- are μ-measurable functions. Then

$$\{x \in X : f(x) > a\} = \begin{cases} \{x \in X : f^+(x) > a\}, & a \geqslant 0, \\ \{x \in X : f^-(x) < -a\}, & a < 0, \end{cases}$$

is a μ-measurable set. This implies that f is a μ-measurable function. □

Theorem 2.13. Assume that $f,g : X \to [-\infty,\infty]$ are μ-measurable functions and $a \in \mathbb{R}$. Then af, $f+g$, $\max(f,g)$, $\min(f,g)$, fg, $\frac{f}{g}$ $(g \neq 0)$, are μ-measurable functions.

W A R N I N G : Since functions are extended real valued, we need to take care about the definitions of $f+g$ and fg. The sum is defined outside the bad set

$$B = \{x \in X : f(x) = \infty \text{ and } g(x) = -\infty\} \cup \{x \in X : f(x) = -\infty \text{ and } g(x) = \infty\},$$

since in B we have $\infty - \infty$ situation. We define

$$(f+g)(x) = \begin{cases} f(x)+g(x), & x \in X \setminus B, \\ a, & x \in B, \end{cases}$$

where $a \in [-\infty,\infty]$ is arbitrary.

Proof. Note that

$$(f+g)^{-1}(\{-\infty\}) = f^{-1}(\{-\infty\}) \cup g^{-1}(\{-\infty\})$$

and

$$(f+g)^{-1}(\{\infty\}) = f^{-1}(\{\infty\}) \cup g^{-1}(\{\infty\})$$

are μ-measurable. Let $a \in \mathbb{R}$. Since

$$\{x \in X : a - f(x) > \lambda\} = \{x \in X : f(x) < a - \lambda\}$$

for every $\lambda \in \mathbb{R}$, the function $a - f$ is μ-measurable. By Lemma 2.10,

$$\{x \in X : f(x) + g(x) > a\} = \{x \in X : g(x) > a - f(x)\}$$

is μ-measurable for every $a \in \mathbb{R}$. This can be also seen directly from

$$(f+g)^{-1}((-\infty,a)) = \bigcup_{r,s \in \mathbb{Q}, r+s<a} \left(f^{-1}((-\infty,r)) \cap g^{-1}((-\infty,s))\right).$$

The functions $|f|$ and g^2 are μ-measurable, see the remark below. Then we may use the formulas

$$\max(f,g) = \frac{1}{2}(f+g+|f-g|), \min(f,g) = \frac{1}{2}(f+g-|f-g|)$$

and

$$fg = \frac{1}{2}((f+g)^2 - f^2 - g^2).$$

□

W A R N I N G : f^2 measurable does not imply that f measurable.

Reason. Let $A \subset X$ be a nonmeasurable set and

$$f : X \to \mathbb{R},\ f(x) = \begin{cases} 1, & x \in A, \\ -1, & x \in X \setminus A. \end{cases}$$

Then $f^2 = 1$ is measurable, but $\{x \in X : f(x) > 0\} = A$ is not a measurable set. ■

2.3 Cantor-Lebesgue function

Recall the construction of the Cantor set from Section 1.11. At the kth step, we have 2^{k-1} open pairwise disjoint open intervals $I_{k,i}$, $i = 1,\dots,2^{k-1}$. Let $m = 1,2,\dots$. Consider all open intervals $I_{k,i}$, with $k = 1,\dots,m$, $i = 1,\dots,2^{k-1}$, used in the construction of the Cantor set at the steps $1,\dots,m$. Note that there are altogether $2^0 + 2^1 + 2^2 + \dots + 2^{m-1} = 2^m - 1$ intervals. Denote these intervals by $\tilde{I}_{m,i}$, $i = 1,\dots,2^m - 1$, organized from left to right.

As in Section 1.11 we have

$$C_0 = [0,1], \quad C_k = [0,1] \setminus \bigcup_{i=1}^{2^k-1} \tilde{I}_{k,i}, \quad k = 1,2,\dots,$$

and the middle thirds Cantor set is $C = \bigcap_{k=0}^{\infty} C_k$. We have seen that C is an uncountable set of Lebesgue measure zero. Define a continuous function $f_k : [0,1] \to [0,1]$ by $f_k(0) = 0$, $f_k(1) = 1$,

$$f_k(x) = \frac{i}{2^k},$$

whenever $x \in \tilde{I}_{k,i}$, $i = 1,2,\dots,2^k - 1$ and f_k is linear on C_k, $k = 1,2,\dots$

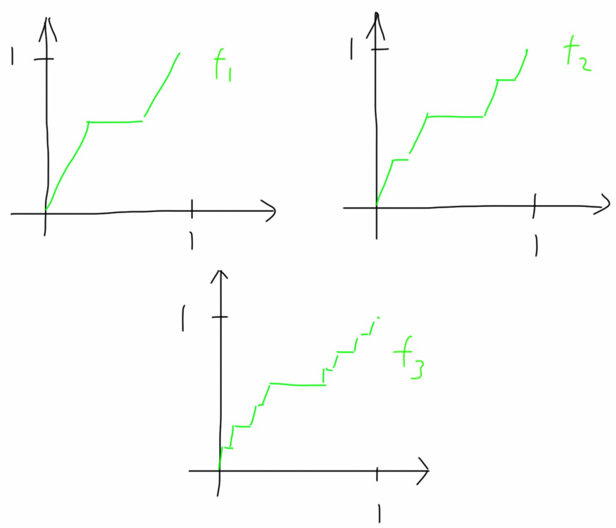

Figure 2.1: The construction of the Cantor-Lebesgue function.

Then $f_k \in C([0,1])$, f is increasing and

$$|f_k(x) - f_{k+1}(x)| < \frac{1}{2^k}$$

for every $x \in [0,1]$. Since

$$|f_k(x) - f_{k+m}(x)| \leqslant \sum_{j=k}^{k+m-1} \frac{1}{2^j} < \frac{1}{2^{k-1}}$$

for every $x \in [0,1]$, (f_k) is a Cauchy sequence in the space $(C([0,1]), \|\cdot\|_\infty)$, where

$$\|f\|_\infty = \sup_{x\in[0,1]} |f(x)|.$$

This is a complete space and thus there exists $f \in C([0,1])$ such that $\|f_k - f\|_\infty \to 0$ as $k \to \infty$. In other words, $f_k \to f$ uniformly in $[0,1]$ as $k \to \infty$. The function f is called the Cantor-Lebesgue function. We collect properties of the Cantor-Lebesgue function below.

(1) $f : [0,1] \to [0,1]$ is continuous, $f(0) = 0$ and $f(1) = 1$.

(2) f is nondecreasing and is constant on each interval in the complement of the Cantor set.

(3) $f : [0,1] \to [0,1]$ is onto, that is, $f([0,1]) = [0,1]$. In fact $f(C) = [0,1]$, that is, for every $y \in [0,1]$ there exists $x \in C$ with $f(x) = y$.

(4) f maps the complement of the Cantor set to a countable set. Thus f maps the Cantor set, which is a set of Lebesgue measure zero, to a set of full measure.

(5) f is locally constant and thus differentiable in the complement of the Cantor set. Thus f is differentiable almost everywhere and its derivative is zero outside the Cantor set. However,

$$\int_{[0,1]} f'(x)\,dx = 0 \neq 1 = f(1) - f(0)$$

and therefore the fundamental theorem of calculus does not hold.

(6) f is not differentiable at any point in the Cantor set.

See [2, p. 67], [7, p. 86-101], [11, p. 38], [12, p. 140–141], [16, p. 87–90] and [15] for more on the Cantor-Lebesgue function.

Remark 2.14. Recall from Section 1.11 that

$$C = \left\{ \sum_{i=1}^{\infty} \frac{a_i}{3^i} : a_i \in \{0,2\},\ i = 1,2,\dots \right\}.$$

It can be shown that if

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} \quad \text{with} \quad a_i \in \{0,2\},\ i = 1,2,$$

then

$$f(x) = \sum_{i=1}^{\infty} \frac{b_i}{2^i} \quad \text{with} \quad b_i = \frac{a_i}{2}.$$

In this sense, the Cantor-Lebesgue function coverts the base three expansions to base two expansions.

Let $g:[0,1]\to[0,2]$, $g(x)=x+f(x)$. Then $g(0)=0$, $g(1)=2$, $g\in C([0,1])$ is strictly increasing and $g([0,1])=[0,2]$. This implies that g is a homeomorphism, that is, g is a continuous function from $[0,1]$ onto $[0,2]$ with a continuous inverse function. Since

$$C=\bigcap_{k=0}^{\infty}C_k=\bigcap_{k=1}^{\infty}\left([0,1]\setminus\bigcup_{i=1}^{2^k-1}\tilde{I}_{k,i}\right)=[0,1]\setminus\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{2^k-1}\tilde{I}_{k,i},$$

where $\tilde{I}_{k,i}$ are pairwise disjoint open intervals,

$$\begin{aligned}
m^*(g(C))&=m^*\left(g([0,1])\setminus\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{2^k-1}g(\tilde{I}_{k,i})\right)\\
&=m^*([0,2])-m^*\left(\bigcup_{k=1}^{\infty}\bigcup_{i=1}^{2^k-1}g(\tilde{I}_{k,i})\right)\quad(g(\tilde{I}_{k,i})\text{ is an interval})\\
&=2-\lim_{k\to\infty}\sum_{i=1}^{2^k-1}m^*(g(\tilde{I}_{k,i}))\quad(g(\tilde{I}_{k,i})\text{ are pairwise disjoint})\\
&=2-\lim_{k\to\infty}\sum_{i=1}^{2^k-1}m^*(\tilde{I}_{k,i})\quad(g(x)=x+a_{k,i}\ \forall\ x\in\tilde{I}_{k,i})\\
&=2-m^*([0,1]\setminus C)=2-1=1.\quad(m^*(C)=0)
\end{aligned}$$

Thus g maps the zero measure Cantor set C to $g(C)$ set of measure one. Since $m(g(C))>0$, by Remark 1.68, there exists $B\subset g(C)$, which is nonmeasurable with respect to the one-dimensional Lebesgue measure. Let $A=g^{-1}(B)$. Then $A\subset C$ and $m(A)=0$. This implies that A is Lebesgue measurable. We collect a few observations related to the Cantor-Lebesgue function below.

(1) The homeomorphism g maps the Cantor set C with $m^*(C)=0$ to a set $g(C)$ with $m^*(g(C))>0$. Sets of Lebesgue measure zero are not mapped to sets of Lebesgue measure zero in continuous mappings.

(2) The homeomorphism g maps a measurable set A to a nonmeasurable set B. Lebesgue measurable sets are not preserved in continuous mappings. Since continuous mappings are Lebesgue measurable, Lebesgue measurable sets are not preserved in Lebesgue measurable mappings.

(3) A is a Lebesgue measurable set that is not a Borel set. Assume for the contradiction, that A is a Borel set. Then $B=g(A)$ is a Borel set, since a homeomorphism maps Borel sets to Borel sets (exercise). However, B is not a Lebesgue measurable set, which implies that it is not a Borel set.

(4) Since $A\subset C$, we conclude that the Cantor set has a subset that is not a Borel set. The set A is Lebesgue measurable subset of the Borel set C with $m^*(C)=0$, but A is not a Borel set. This shows that the restriction of the Lebesgue measure to the Borel sets is not complete.

(5) $\chi_A\circ g^{-1}=\chi_B$, where the function χ_B is nonmeasurable, but the functions χ_A and g^{-1} are measurable functions, since A is a measurable set and g^{-1}

is continuous. A composed function of two Lebesgue measurable functions is not Lebesgue measurable. For positive results, see Lemma 2.7.

(6) g^{-1} is a measurable function which does not satisfy that $(g^{-1})^{-1}(A)$ is measurable for every measurable set A.

2.4 Lipschitz mappings on $\mathbb{R}^n$

The Cantor-Lebesgue function showed that Lebesgue measurability of a set is not necessarily preserved in continuous mappings. In this section we study certain conditions for a function $f:\mathbb{R}^n \to \mathbb{R}^n$, which guarantee that f maps Lebesgue measurable sets to Lebesgue measurable sets.

Definition 2.15. A mapping $f:\mathbb{R}^n \to \mathbb{R}^n$ is said to be Lipschitz continuous, if there exists a constant L such that

$$|f(x)-f(y)| \leqslant L|x-y|$$

for every $x, y \in \mathbb{R}^n$.

Remark 2.16. A mapping $f:\mathbb{R}^n \to \mathbb{R}^n$ is of the form $f(x) = (f_1(x),\dots,f_n(x))$, where $x = (x_1,\dots,x_n)$ and the coordinate functions $f_i:\mathbb{R}^n \to \mathbb{R}$ for $i = 1,\dots,n$. Such a mapping f is Lipschitz continuous if and only if all coordinate functions f_i, $i = 1,\dots,n$, satisfy a Lipschitz condition

$$|f_i(x)-f_i(y)| \leqslant L_i|x-y|$$

for every $x, y \in \mathbb{R}^n$ with some constant L_i.

Examples 2.17:

(1) Every linear mapping $L:\mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous.

Reason. Let A be the $n\times n$-matrix representing L. Then

$$|L(x)-L(y)| = |Ax-Ay| = |A(x-y)| \leqslant \|A\||x-y|$$

for every $x, y \in \mathbb{R}^n$, where $\|A\| = \max\{|a_{ij}| : i, j = 1,\dots,n\}$. ■

(2) Every mapping $f:\mathbb{R}^n \to \mathbb{R}^n$, $f = (f_1,\dots,f_n)$, whose coordinate functions f_i, $i = 1,\dots,n$, have bounded first partial derivatives in $\mathbb{R}^n$, is Lipschitz continuous.

Reason. By the fundamental theorem of calculus,

$$f_i(x)-f_i(y) = \int_0^1 \frac{\partial}{\partial t}(f_i((1-t)x+ty))\,dt = \int_0^1 \nabla f_i((1-t)x+ty)\cdot(y-x)\,dt.$$

This implies

$$|f_i(x)-f_i(y)| \leqslant \int_0^1 |\nabla f_i((1-t)x+ty)||x-y|\,dt \leqslant \sup_{z\in\mathbb{R}^n}|\nabla f_i(z)||x-y|$$

for every $x, y \in \mathbb{R}^n$. ■

Lemma 2.18. Assume that $f:\mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz continuous mapping. Then $m^*(f(A)) = 0$ whenever $m^*(A) = 0$.

THE MORAL: A Lipschitz mapping $f:\mathbb{R}^n \to \mathbb{R}^n$ maps sets of Lebesgue measure zero to sets of Lebesgue measure zero.

Proof. Assume that $m^*(A) = 0$ and let $\varepsilon > 0$. Then (exercise) there are balls $\{B(x_i,r_i)\}_{i=1}^{\infty}$ such that

$$A \subset \bigcup_{i=1}^{\infty} B(x_i,r_i) \quad \text{and} \quad \sum_{i=1}^{\infty} m^*(B(x_i,r_i)) < \varepsilon.$$

By the Lipschitz condition,

$$|f(x_i) - f(y)| \leqslant L|x_i - y| < Lr_i,$$

for every $y \in B(x_i,r_i)$ and thus $f(y) \in B(f(x_i), Lr_i)$. This implies

$$f(A) \subset f\Big(\bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) = \bigcup_{i=1}^{\infty} f(B(x_i,r_i)) \subset \bigcup_{i=1}^{\infty} B(f(x_i),Lr_i).$$

By monotonicity, countable subadditivity translation invariance and the scaling property of the Lebesgue measure we have

$$\begin{aligned} m^*(f(A)) &\leqslant m^*\Big(\bigcup_{i=1}^{\infty} B(f(x_i),Lr_i)\Big) \leqslant \sum_{i=1}^{\infty} m^*\big(B(f(x_i),Lr_i)\big) \\ &= L^n \sum_{i=1}^{\infty} m^*(B(x_i,r_i)) < L^n \varepsilon. \end{aligned}$$

This implies that $m^*(f(A)) = 0$. □

Theorem 2.19. Assume that $f:\mathbb{R}^n \to \mathbb{R}^n$ is a continuous function which maps sets of Lebesgue measure zero to sets of Lebesgue measure zero. Then f maps Lebesgue measurable sets to Lebesgue measurable sets.

THE MORAL: In particular, a Lipschitz mapping $f:\mathbb{R}^n \to \mathbb{R}^n$ maps Lebesgue measurable sets to Lebesgue measurable sets.

Proof. Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set. Consider $A_k = A \cap B(0,k)$ with $m^*(A_k) < \infty$ for every $k = 1,2,\dots.$ By Theorem 1.59 there exist compact sets $K_i \subset A_k$, $i = 1,2,\dots$, such that

$$m^*\big(A_k \setminus \bigcup_{i=1}^{\infty} K_i\big) = 0.$$

Since the set A_k can be written as

$$A_k = \Big(\bigcup_{i=1}^{\infty} K_i\Big) \cup \Big(A_k \setminus \bigcup_{i=1}^{\infty} K_i\Big),$$

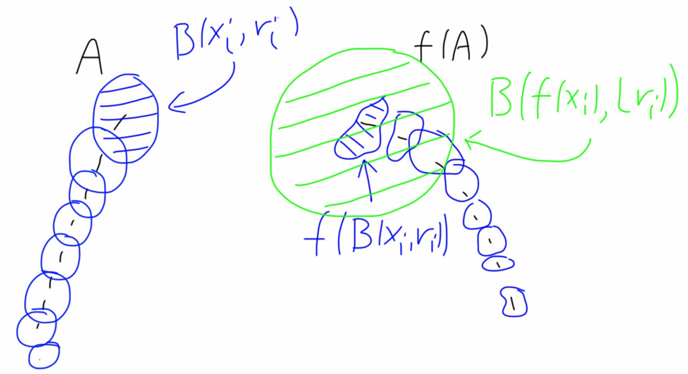

Figure 2.2: The image of a set of measure zero.

we have

$$f(A_k) = \bigcup_{i=1}^{\infty} f(K_i) \cup f\big(A_k \setminus \bigcup_{i=1}^{\infty} K_i\big).$$

Since a continuous function maps compact set to compact sets and compact sets are Lebesgue measurable, the countable union $\bigcup_{i=1}^{\infty} f(K_i)$ is a Lebesgue measurable set. On the other hand, by Lemma 2.18 the function f maps sets of measure zero to sets of measure zero. This implies that $f\big(A_k \setminus \bigcup_{i=1}^{\infty} K_i\big)$ is of measure zero and thus Lebesgue measurable. The set $f(A_k)$ is measurable as a union of two measurable sets. Finally

$$f(A) = f\left(\bigcup_{k=1}^{\infty} A_k\right) = \bigcup_{k=1}^{\infty} f(A_k)$$

is Lebesgue measurable as a countable union of Lebesgue measurable sets. □

Remark 2.20. If $f : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz mapping with constant L, then there exists a constant c, depending only on L and n, such that

$$m^*(f(A)) \leqslant c m^*(A)$$

for every set $A \subset \mathbb{R}^n$ (exercise).

WARNING: It is important that the source and the target dimensions are same in the results above. There is a measurable subset A of $\mathbb{R}^n$ with respect to the n-dimensional Lebesgue measure such that the projection to the first coordinate axis is not Lebesgue measurable with respect to the one-dimensional Lebesgue measure. Observe that the projection is a Lipschitz continuous mapping from $\mathbb{R}^n$ to $\mathbb{R}$ with the Lipschitz constant one, see Remark 1.69.

2.5 Limits of measurable functions

Next we show that measurability is preserved under limit operations.

Theorem 2.21. Assume that $f_i : X \to [-\infty,\infty]$, $i = 1,2,\dots$, are μ-measurable functions. Then

$$\sup_i f_i, \quad \inf_i f_i, \quad \limsup_{i\to\infty} f_i \quad \text{and} \quad \liminf_{i\to\infty} f_i$$

are μ-measurable functions.

Remark 2.22. Recall that

$$\limsup_{i\to\infty} f_i(x) = \inf_{j\geqslant 1}(\sup_{i\geqslant j} f_i(x))$$

and

$$\liminf_{i\to\infty} f_i(x) = \sup_{j\geqslant 1}(\inf_{i\geqslant j} f_i(x)).$$

Proof. Since

$$\{x \in X : \sup_i f_i(x) > a\} = \bigcup_{i=1}^{\infty} \{x \in X : f_i(x) > a\}$$

for every $a \in \mathbb{R}$, the function $\sup_i f_i$ is μ-measurable. The measurability of $\inf_i f_i$ follows from

$$\inf_i f_i(x) = -\sup_i(-f_i(x))$$

or from

$$\{x \in X : \inf_i f_i(x) < a\} = \bigcup_{i=1}^{\infty} \{x \in X : f_i(x) < a\}$$

for every $a \in \mathbb{R}$. The claims that $\limsup_{i\to\infty} f_i$ and $\liminf_{i\to\infty} f_i$ are μ-measurable functions follow immediately. □

Theorem 2.23. Assume that $f_i : X \to [-\infty,\infty]$, $i = 1,2,\dots$, are μ-measurable functions such that the sequence $(f_i(x))$ converges for every $x \in X$ as $i \to \infty$. Then

$$f = \lim_{i\to\infty} f_i$$

is a μ-measurable function.

THE MORAL: Measurability is preserved in taking limits. This is a very important property of a measurable function.

Proof. This follows from the previous theorem, since

$$f = \limsup_{i\to\infty} f_i = \liminf_{i\to\infty} f_i$$

□

Example 2.24. Assume that $f : \mathbb{R} \to \mathbb{R}$ is differentiable. Then f is continuous and thus Lebesgue measurable. Morever, the difference quotients

$$g_i(x) = \frac{f(x+\frac{1}{i}) - f(x)}{\frac{1}{i}}, \quad i = 1, 2, \ldots,$$

are continuous and thus Lebesgue measurable. Hence

$$f' = \lim_{i\to\infty} g_i$$

is a Lebesgue measurable function. Note that f' is not necessarily continuous.

Reason. The function $f : \mathbb{R} \to \mathbb{R}$

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

is differentiable everywhere, but f' is not continuous at $x = 0$. ■

2.6 Almost everywhere

Definition 2.25. Let μ be an outer measure in X. A property is said to hold μ-almost everywhere in X, if it holds in $X \setminus A$ for a set $A \subset X$ with $\mu(A) = 0$. It is sometimes denoted that that the property holds μ-a.e.

THE MORAL: Sets of measure zero are negligible sets in the measure theory. In other words, a measure does not see sets of measure zero. Measure theory is very flexible, since sets of measure zero do not affect the measurability of a set or a function. The price we have to pay is that we can obtain information only up to sets of measure zero by measure theoretical tools.

Remark 2.26. Almost everywhere is called "almost surely" in probability theory.

Examples 2.27:

(1) The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \chi_{(0,\infty)}(x)$, is continuous almost everywhere, because the set of discontinuity $\{0\}$ has Lebesgue measure zero.

(2) The function $x \mapsto |x|$ is almost everyhwere differentiable, because the set of non-differentiable points $\{0\}$ has Lebesgue measure zero.

(3) Many useful functions such as

$$f:\mathbb{R}\to\mathbb{R},\, f(x)=\frac{\sin(x)}{x} \quad \text{and} \quad f:\mathbb{R}^n\to\mathbb{R},\, f(x)=|x|^{-\alpha},\, \alpha>0,$$

are defined only almost everywhere with respect to the Lebesgue measure.

(4) Let $\alpha>0$. The function $f:\mathbb{R}^n\to[-\infty,\infty]$,

$$f(x)=\begin{cases}|x|^{-\alpha}, & x\neq 0,\\ \infty, & x=0,\end{cases}$$

is finite almost everywhere, because the infinity set $f^{-1}(\{\infty\})=\{0\}$ has Lebesgue measure zero.

Lemma 2.28. Assume that $f:X\to[-\infty,\infty]$ is a μ-measurable function. If $g:X\to[-\infty,\infty]$ is a function with $f=g$ μ-almost everywhere in X, then g is a μ-measurable function.

THE MORAL: In case $f=g$ μ-almost everywhere, we do not usually distinguish f from g. Measure theoretically they are the same function. To be very formal, we could define an equivalence relation

$$f\sim g \iff f=g \quad \mu\text{-almost everywhere},$$

but this is hardly necessary.

Proof. Let $A=\{x\in X: f(x)\neq g(x)\}$. By assumption $\mu(A)=0$. Then

$$\begin{aligned}\{x\in X: g(x)>a\}&=(\{x\in X: g(x)>a\}\cap A)\cup(\{x\in X: g(x)>a\}\cap(X\setminus A))\\&=(\{x\in X: g(x)>a\}\cap A)\cup(\{x\in X: f(x)>a\}\cap(X\setminus A))\end{aligned}$$

is a μ-measurable set for every $a\in\mathbb{R}$, since $\mu((\{x\in X: g(x)>a\}\cap A))=0$. □

Remark 2.29. All properties of measurable functions can be relaxed to conditions holding almost everywhere. For example, if $f_i:X\to[-\infty,\infty]$, $i=1,2,\dots$, are μ-measurable functions and

$$f=\lim_{i\to\infty} f_i$$

μ-almost everywhere, then f is a μ-measurable function. Moreover, if the functions f and g are defined almost everywhere, the functions $f+g$ and fg are defined only in the intersection of the domains of f and g. Since the union of two sets of measure zero is a set of measure zero the functions are defined almost everywhere.

Remark 2.30. We discuss property that holds almost everywhere on a measure space $(X,\mathcal{M},\mu)$, see Definition 1.11. A property of points of X is said to hold μ-almost everywhere in X, if there exists a set $A\in\mathcal{M}$ with $\mu(A)=0$, such that

A contains every point at which the property does to hold. Consider a property that holds μ-almost everywhere, and let B be the set of points in X at which it does not hold. Then it is not necessary that $B \in \mathscr{A}$, but that there exists a set $A \in \mathscr{A}$ with $B \subset A$ and $\mu(A) = 0$. If μ is a complete measure, then $B \in \mathscr{A}$. See Definition 1.13. Let $(X, \mathscr{M}, \mu)$ be a measure space that is not complete, let $A \in \mathscr{M}$ be a set with $\mu(A) = 0$ and let $B \subset A$ be a set with $B \notin \mathscr{M}$. Then $f = 0$ and $g = \chi_B$ satisfy $f = g$ in $X \setminus A$ and thus $f = g$ μ-almost everywhere in X. However, f is a measurable function, but g is not. Thus Lemma 2.28 does not hold for measures that are not complete. In addition, the sequence $f_i = 0$ converges to g μ-almost everywhere as $i \to \infty$, so that the limit of measurable functions that converge μ-almost everywhere is not necessarily μ-meausurable. Recall that all outer measures are complete by Remark 1.5 (3) and these problems do not occur.

2.7 Approximation by simple functions

Next we consider the approximation of a measurable function with simple functions, which are the basic blocks in the definition of the integral.

Definition 2.31. A function $f : X \to \mathbb{R}$ is simple, if its range is a finite set $\{a_1, \dots, a_n\}$, $n \in \mathbb{N}$, and the preimages

$$f^{-1}(\{a_i\}) = \{x \in X : f(x) = a_i\}$$

are μ-measurable sets.

T H E M O R A L : A simple function is a linear combination of finitely many characteristic functions of measurable sets, since it can be uniquely written as a finite sum

$$f = \sum_{i=1}^{n} a_i \chi_{A_i},$$

where $A_i = f^{-1}(\{a_i\})$. Remark 2.2 (3) and Theorem 2.13 imply that a simple function is μ-measurable.

W A R N I N G : A simple function assumes only finitely many values, but the sets $A_i = f^{-1}(\{a_i\})$ may not be geometrically simple.

Reason. The function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \chi_{\mathbb{Q}}(x)$ is simple with respect to the one dimensional Lebesgue measure, but it is discontinuous at every point. In particular, it is possible that a measurable function is discontinuous at every point and thus it does not have any regularity in this sense. ■

We consider approximation properties of measurable functions.

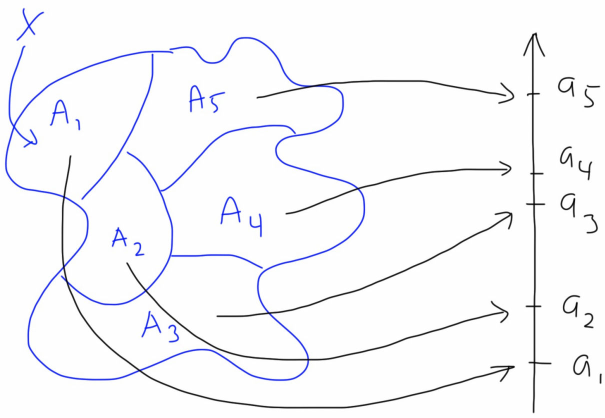

Figure 2.3: A simple function.

Theorem 2.32. Assume that $f : X \to [0,\infty]$ is a nonnegative function. Then f is a μ-measurable function if and only if there exists an increasing sequence f_i, $i = 1,2,\dots,$ of simple functions such that

$$f(x) = \lim_{i\to\infty} f_i(x)$$

for every $x \in X$.

T H E M O R A L : Every nonnegative measurable function can be approximated by an increasing sequence of simple functions.

Proof. $\boxed{\Longrightarrow}$ For every $i = 1,2,\dots$ partition $[0,i)$ into $i2^i$ intervals

$$I_{i,k} = \left[\frac{k-1}{2^i}, \frac{k}{2^i}\right), \quad k = 1,\dots,i2^i.$$

Denote

$$A_{i,k} = f^{-1}(I_{i,k}) = \left\{x \in X : \frac{k-1}{2^i} \leqslant f(x) < \frac{k}{2^i}\right\}, \quad k = 1,\dots,i2^i,$$

and

$$A_i = f^{-1}([i,\infty]) = \{x \in X : f(x) \geqslant i\}.$$

These sets are μ-measurable and they form a pairwise disjoint partition of X. The approximating simple function is defined as

$$f_i(x) = \sum_{k=1}^{i2^i} \frac{k-1}{2^i}\chi_{A_{i,k}}(x) + i\chi_{A_i}(x).$$

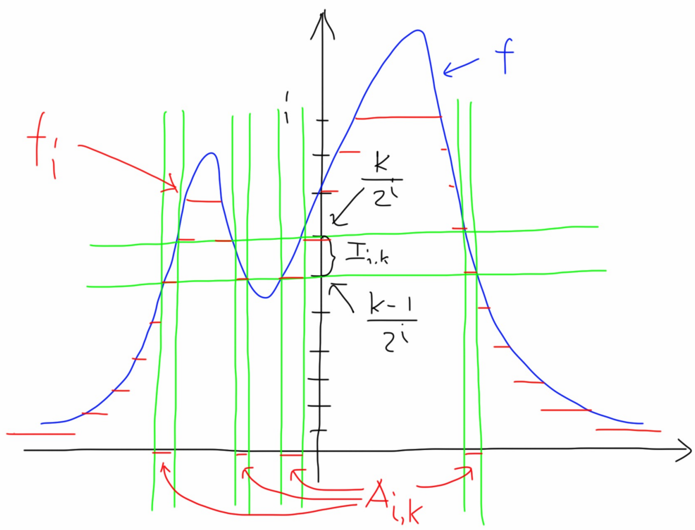

Figure 2.4: Approximation by simple functions.

Since the sets are pairwise disjoint, $0 \leqslant f_i(x) \leqslant f_{i+1}(x) \leqslant f(x)$ for every $x \in X$. In addition,

$$|f(x) - f_i(x)| \leqslant \frac{1}{2^i}, \quad \text{if} \quad x \in \bigcup_{k=1}^{i2^i} A_{i,k} = \{x \in X : f(x) < i\}$$

and

$$f_i(x) = i, \quad \text{if} \quad x \in A_i = \{x \in X : f(x) \geqslant i\}.$$

This implies that

$$\lim_{i\to\infty} f_i(x) = f(x)$$

for every $x \in X$.

$\boxed{\Longleftarrow}$ Follows from the fact that a poitwise limit of measurable functions is measurable, see Theorem 2.23. □

Remark 2.33. As the proof above shows, the approximation by simple functions is based on a subdivision of the range instead of the domain, as in the case of step functions. The approximation procedure is compatible with the definition of a measurable function.

Next we consider sign-changing functions.

Corollary 2.34. The function $f : X \to [-\infty, \infty]$ is a μ-measurable function if and only if there exist simple functions f_i, $i = 1, 2, \ldots$, such that

$$f(x) = \lim_{i\to\infty} f_i(x)$$

for every $x \in X$.

THE MORAL: A function is measurable if and only if it can be approximated pointwise by simple functions.

Proof. We use the decomposition $f = f^+ - f^-$. By Theorem 2.32 there are simple functions g_i and h_i, $i = 1, 2, \dots$, such that

$$f^+ = \lim_{i\to\infty} g_i \quad \text{and} \quad f^- = \lim_{i\to\infty} h_i.$$

The functions $f_i = g_i - h_i$ do as an approximation. □

Remarks 2.35:

(1) The sequence $(|f_i|)$ is increasing, that is, $|f_i| \leqslant |f_{i+1}| \leqslant |f|$ for every $i = 1, 2, \dots$, because $|f_i| = g_i + h_i$ and the sequences (g_i) and (h_i) are increasing.

(2) If the limit function f is bounded, then the simple functions will converge uniformly to f in X.

(3) This approximation holds for every function f, but in that case the simple functions are not measurable.

2.8 Modes of convergence

Let us recall two classical modes for a sequence of functions $f_i : X \to \mathbb{R}$, $i = 1, 2, \dots$, to converge to a function $f : X \to \mathbb{R}$.

(1) f_i converges pointwise to f, if

$$|f_i(x) - f(x)| \to 0 \quad \text{as} \quad i \to \infty$$

for every $x \in X$. This means that for every $\varepsilon > 0$, there exists i_ε such that

$$|f_i(x) - f(x)| < \varepsilon$$

whenever $i \geqslant i_\varepsilon$. Note that i_ε depends on x and ε.

(2) f_i converges uniformly to f in X, if

$$\sup_{x\in X} |f_i(x) - f(x)| \to 0 \quad \text{as} \quad i \to \infty.$$

This means that for every $\varepsilon > 0$, there exists i_ε such that

$$|f_i(x) - f(x)| < \varepsilon$$

for every $x \in X$ whenever $i \geqslant i_\varepsilon$. In this case i_ε does not depend on x.

Example 2.36. A uniform convergence implies pointwise convergence, but the converse is not true. For example, let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \frac{x}{i}, \quad i = 1, 2, \ldots$$

Then $f_i \to 0$ pointwise, but not uniformly in $\mathbb{R}$.

There are also other modes of convergence that are relevant in measure theory. For simplicity we will discuss only real-valued functions. Everything can be extended to the case when $f_i : X \to [-\infty, \infty]$ and $f : X \to [-\infty, \infty]$ are μ-measurable functions with $|f_i| < \infty$ and $|f| < \infty$ μ-almost everywhere in X for every $i = 1, 2, \ldots$. In this case we consider a set $A \subset X$ with $\mu(A) = 0$ such that $|f_i|$, $i = 1, 2, \ldots$, and $|f|$ are finite in the complement of A and replace f_i and f by $g_i = f_i \chi_{X \setminus A}$ and $g = \chi_{X \setminus A} g$. This allows us to avoid expressions as $|f_i(x) - f(x)|$ when $|f_i(x)|$ or $|f(x)|$ is infinite.

Definition 2.37 (Convergence almost everywhere). We say that f_i converges to f almost everywhere in X, if $f_i(x) \to f(x)$ for μ-almost every $x \in X$.

T H E M O R A L : Almost everywhere convergence is pointwise convergence outside a set of measure zero.

Remark 2.38. $f_i \to f$ almost everywhere in X if and only if for every $\varepsilon > 0$ there exists a μ-measurable set $A \subset X$ such that $\mu(X \setminus A) < \varepsilon$ and $f_i \to f$ pointwise in A.

Reason. $\boxed{\Longrightarrow}$ If $f_i \to f$ almost everywhere in X, there exists a set A with $\mu(A) = 0$ such that $f_i(x) \to f(x)$ for every $x \in X \setminus A$. The set A satisfies the required properties.

$\boxed{\Longleftarrow}$ Assume that for every $j = 1, 2, \ldots$ there exists a μ-measurable set $A_j \subset X$ such that $\mu(X \setminus A_j) < \frac{1}{j}$ and $f_i(x) \to f(x)$ for every $x \in A_j$ as $i \to \infty$. Consider $A = \bigcup_{j=1}^{\infty} A_j$. Then

$$\mu(X \setminus A) = \mu\left(X \setminus \bigcup_{j=1}^{\infty} A_j\right) = \mu\left(\bigcap_{j=1}^{\infty} (X \setminus A_j)\right) \leqslant \mu(X \setminus A_j) < \frac{1}{j}$$

for every $j = 1, 2, \ldots$. This implies $\mu(X \setminus A) = 0$. Let $x \in A = \bigcup_{j=1}^{\infty} A_j$. Then $x \in A_j$ for some j and $f_i(x) \to f(x)$ as $i \to \infty$. This shows that $f_i(x) \to f(x)$ for every $x \in A$ as $i \to \infty$. ■

Definition 2.39 (Almost uniform convergence). We say that f_i converges to f almost uniformly in X, if for every $\varepsilon > 0$ there is a μ-measurable set $A \subset X$ such that $\mu(X \setminus A) < \varepsilon$ and $f_i \to f$ uniformly in A.

T H E M O R A L : Almost uniform convergence is uniform convergence outside a set of arbitrarily small measure.

Remark 2.40. If $f_i \to f$ almost uniformly in X, then $f_i \to f$ almost everywhere in X, see Remark 2.38.

Example 2.41. Let $f_i : [0,1] \to \mathbb{R}$, $f_i(x) = x^i$, $i = 1,2,\dots$ and

$$f(x) = \begin{cases} 0, & 0 \leqslant x < 1, \\ 1, & x = 1. \end{cases}$$

Then $f_i(x) \to f(x)$ for every $x \in [0,1]$, but f_i does not converge to f uniformly in $[0,1]$. However, $f_i \to f$ almost uniformly in $[0,1]$, since $f_i \to f$ uniformly in every $[0,1-\varepsilon]$ with $0 < \varepsilon < \frac{1}{2}$. This example also shows that almost uniform convergence does not imply uniform convergence outside a set of measure zero.

Definition 2.42 (Convergence in measure). We say that f_i converges to f in measure in X, if

$$\lim_{i\to\infty} \mu(\{x \in X : |f_i(x) - f(x)| \geqslant \varepsilon\}) = 0$$

for every $\varepsilon > 0$.

THE MORAL: It is instructive to compare almost uniform convergence with convergence in measure. Assume that $f_i \to f$ in measure on X. Let $\varepsilon > 0$. Then there exists a set A_i such that $|f_i(x) - f(x)| < \varepsilon$ for every $x \in A_i$ with $\mu(X \setminus A_i) < \varepsilon$. Note that the sets A_i may vary with i, as in the case of a sliding sequence of functions above. Almost uniform convergence requires that a single set A will do for all sufficiently large indices, that is, the set A does not depend on i.

Remark 2.43. If $f_i \to f$ almost uniformly in X, then $f_i \to f$ in measure in X.

Reason. For $j = 1,2,\dots$ there exists a μ-measurable set A_j such that $\mu(X \setminus A_j) < \frac{1}{j}$ and $f_i \to f$ uniformly in A_j. Let $\varepsilon > 0$. There exists i_ε such that

$$\sup_{x\in A_j} |f_i(x) - f(x)| < \varepsilon,$$

whenever $i \geqslant i_\varepsilon$. This implies

$$\mu(\{x \in X : |f_i(x) - f(x)| \geqslant \varepsilon\}) \leqslant \mu(X \setminus A_j) < \frac{1}{j},$$

whenever $i \geqslant i_\varepsilon$. It follows that

$$\limsup_{i\to\infty} \mu(\{x \in X : |f_i(x) - f(x)| \geqslant \varepsilon\}) \leqslant \frac{1}{j}.$$

By letting $j \to \infty$, we obtain

$$\lim_{i\to\infty} \mu(\{x \in X : |f_i(x) - f(x)| \geqslant \varepsilon\}) = 0.$$

■

Next we give examples which distinguish between the modes of convergence. In the following moving bump examples we have $X = \mathbb{R}$ with the Lebesgue measure.

Examples 2.44:

(1) (Escape to horizontal infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \chi_{[i,i+1]}(x), \quad i = 1,2,\dots$$

Then $f_i \to 0$ everywhere and thus almost everywhere in $\mathbb{R}$, but not uniformly, almost uniformly or in measure.

(2) (Escape to width infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \frac{1}{i}\chi_{[0,i]}(x), \quad i = 1,2,\dots$$

Then $f_i \to 0$ uniformly in $\mathbb{R}$.

(3) (Escape to vertical infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = i\chi_{\left[\frac{1}{i},\frac{2}{i}\right]}(x), \quad i = 1,2,\dots$$

Then $f_i \to 0$ pointwise, almost uniformly and in measure, but not uniformly in $\mathbb{R}$.

(4) (A sliding sequence of functions) Let $f_i : [0,1] \to \mathbb{R}$, $i = 1,2,\dots$, be defined by

$$f_{2^k+j}(x) = k\chi_{\left[\frac{j}{2^k},\frac{j+1}{2^k}\right]}(x), \quad k = 0,1,2,\dots, \quad j = 0,1,\dots,2^k - 1.$$

Then

$$\limsup_{i\to\infty} f_i(x) = \infty \quad \text{and} \quad \liminf_{i\to\infty} f_i(x) = 0$$

for every $x \in [0,1]$ and thus the pointwise limit does not exist at any point. However,

$$m^*(\{x \in [0,1] : f_{2^k+j}(x) \geqslant \varepsilon\}) = m^*\left(\left[\frac{j}{2^k},\frac{j+1}{2^k}\right]\right) = \frac{1}{2^k} \to 0 \quad \text{as} \quad k \to \infty.$$

This shows that $f_i \to 0$ in measure on $[0,1]$. Note that there are several converging subsequences. For example, $f_{2^i}(x) \to 0$ for every $x \neq 0$, although the original sequence diverges everywhere.

The next result shows that, for a sequence that converges in measure, a converging subsequence, as in the sliding sequence of functions above, always exists.

Theorem 2.45. Assume that $f_i \to f$ in measure. There exists a subsequence (f_{i_k}) such that $f_{i_k} \to f$ μ-almost everywhere.

Proof. Choose i_1 such that

$$\mu(\{x \in X : |f_{i_1}(x) - f(x)| \geqslant 1\}) < \frac{1}{2}.$$

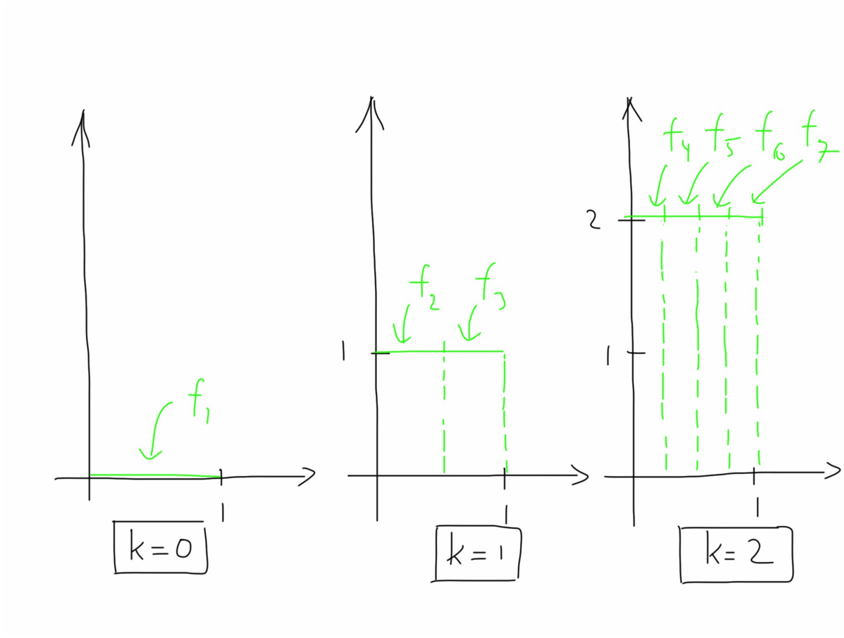

Figure 2.5: A sliding sequence of functions.

Assume then that $i_1,\dots,i_k$ have been chosen. Choose $i_{k+1} > i_k$ such that

$$\mu\left(\left\{x \in X : |f_{i_{k+1}}(x) - f(x)| \geqslant \frac{1}{k+1}\right\}\right) < \frac{1}{2^{k+1}}.$$

Define

$$A_j = \bigcup_{k=j}^{\infty} \left\{x \in X : |f_{i_k}(x) - f(x)| \geqslant \frac{1}{k}\right\}, \quad j = 1,2,\dots$$

Clearly $A_{j+1} \subset A_j$ and denote $A = \bigcap_{j=1}^{\infty} A_j$. For every $j = 1,2,\dots$, we have

$$\mu(A) \leqslant \mu(A_j) \leqslant \sum_{k=j}^{\infty} \frac{1}{2^k} = \frac{2}{2^j}$$

and by letting $j \to \infty$ we conclude $\mu(A) = 0$. By de Morgan's law

$$\begin{aligned}
X \setminus A = X \setminus \bigcap_{j=1}^{\infty} A_j &= \bigcup_{j=1}^{\infty} (X \setminus A_j) \\
&= \bigcup_{j=1}^{\infty} \left(X \setminus \bigcup_{k=j}^{\infty} \left\{x \in X : |f_{i_k}(x) - f(x)| \geqslant \frac{1}{k}\right\}\right) \\
&= \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} \left(X \setminus \left\{x \in X : |f_{i_k}(x) - f(x)| \geqslant \frac{1}{k}\right\}\right) \\
&= \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} \left\{x \in X : |f_{i_k}(x) - f(x)| < \frac{1}{k}\right\}.
\end{aligned}$$

For every $x \in X \setminus A$ there exists j such that

$$|f_{i_k}(x) - f(x)| < \frac{1}{k} \quad \text{for every} \quad k \geqslant j.$$

This implies that $|f_{i_k}(x) - f(x)| \to 0$ for every $x \in X \setminus A$ as $k \to \infty$. □

Remark 2.46. The following assertions are valid for almost everywhere convergence, almost uniform convergence and convergence in measure.

(1) The limit function f is μ-measurable. See Theorem 2.23 and the discussion in Section 2.6.

(2) The limit function f is unique up to a set of μ-measure zero.

(3) Convergence is not affected by changing f_i or f on a set of μ-measure zero.

Remark 2.47. Convergence almost everywhere is called "convergence almost surely" in probability theory and convergence in measure is called "convergence in probability".

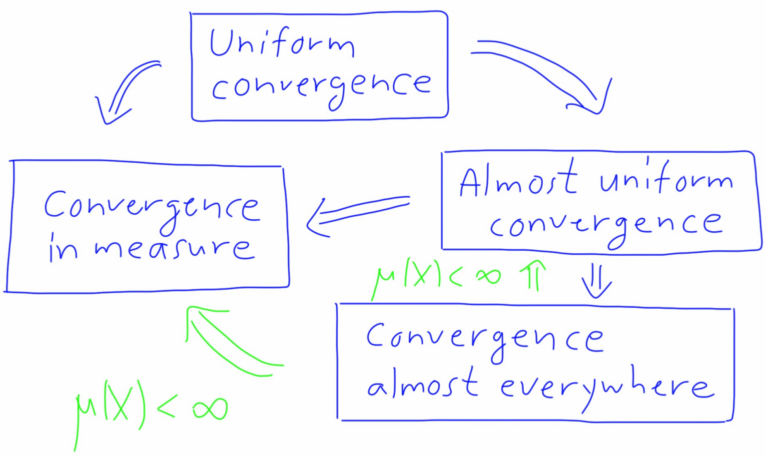

Figure 2.6: Comparison of modes of convergence.

2.9 Egoroff's and Lusin's theorems

The next result gives the main motivation for almost uniform convergence.

Theorem 2.48 (Egoroff's theorem). Assume that $\mu(X) < \infty$. Let $f_i : X \to [-\infty, \infty]$ be a μ-measurable function with $|f_i| < \infty$ μ-almost everywhere for every $i = 1, 2, \ldots$ such that $f_i \to f$ almost everywhere in X and $|f| < \infty$ μ-almost everywhere. Then $f_i \to f$ almost uniformly in X.

THE MORAL: Almost uniform convergence and almost everywhere convergence are equivalent in a space with finite measure.

Remark 2.49. The sequence $f_i = \chi_{B(0,i)}$, $i = 1,2,\dots$, converges pointwise to $f = 1$ in $\mathbb{R}^n$, but it does not converge uniformly outside any bounded set. On the other hand, if $|f_i| < \infty$ everywhere, but $|f| = \infty$ on a set of positive measure, then $|f_i - f| = \infty$ on a set of positive measure. Hence these assumptions cannot be removed.

Proof. Let $\varepsilon > 0$. Define

$$A_{j,k} = \bigcup_{i=j}^{\infty} \left\{ x \in X : |f_i(x) - f(x)| > \frac{1}{2^k} \right\}, \quad j,k = 1,2,\dots$$

Then $A_{j+1,k} \subset A_{j,k}$ for every $j,k = 1,2,\dots$. Since $\mu(X) < \infty$ and $f_i \to f$ almost everywhere in X, we have

$$\lim_{j\to\infty} \mu(A_{j,k}) = \mu\left(\bigcap_{j=1}^{\infty} A_{j,k} \right) = 0.$$

Thus there exists j_k such that

$$\mu(A_{j_k,k}) < \frac{\varepsilon}{2^{k+1}}.$$

Denote $A = X \setminus \bigcup_{k=1}^{\infty} A_{j_k,k}$. This implies

$$\mu(X \setminus A) \leqslant \sum_{k=1}^{\infty} \mu(A_{j_k,k}) \leqslant \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+1}} < \varepsilon.$$

Then for every $k = 1,2,\dots$ and $i \geqslant j_k$ we have

$$|f_i(x) - f(x)| \leqslant \frac{1}{2^k}$$

for every $x \in A$. This implies that $f_i \to f$ uniformly in A. □

Remark 2.50. If $\mu(X) = \infty$, we can apply Egoroff's theorem for μ-measurable subsets $A \subset X$ with $\mu(A) < \infty$. As far as $\mathbb{R}^n$ is concerned, a sequence (f_i) is said to converge locally uniformly to f, if $f_i \to f$ uniformly on every bounded set $A \subset \mathbb{R}^n$. Equivalently, we could require that for every point $x \in \mathbb{R}^n$ there is a ball $B(x,r)$, with $r > 0$, such that $f_i \to f$ uniformly in $B(x,r)$. Let us rephrase Egoroff's theorem for the Lebesgue measure, or a more general Radon measure, on $\mathbb{R}^n$. Let (f_i) be a sequence measurable functions with $|f_i| < \infty$ almost everywhere for every $i = 1,2,\dots$ such that $f_i \to f$ almost everywhere in $\mathbb{R}^n$ and $|f| < \infty$ almost everywhere. Then for every $\varepsilon > 0$ there exists a measurable set $A \subset \mathbb{R}^n$ such that the measure of A is at most ε and $f_i \to f$ locally uniformly in $\mathbb{R}^n \setminus A$.

Remark 2.51. Relations of measurable sets and functions to standard open sets and continuous functions are summarized in Littlewood's three principles.

(1) Every measurable set is almost open (Theorem 1.44).

(2) Pointwise convergence is almost uniform (Egoroff's theorem 2.48).

(3) A measurable function is almost continuous (Lusin's theorem 2.52).

Here the word "almost" has to be understood measure theoretically.

The following result is related to Littlewood's third principle. We shall prove it only in the case $X = \mathbb{R}^n$, but the result also holds in more general metric spaces with the same proof.

Theorem 2.52 (Lusin's theorem). Let μ be a Borel regular outer measure on $\mathbb{R}^n$, $A \subset \mathbb{R}^n$ a μ-measurable set such that $\mu(A) < \infty$ and $f : \mathbb{R}^n \to [-\infty, \infty]$ be a μ-measurable function such that $|f| < \infty$ μ-almost everywhere. For every $\varepsilon > 0$ there exists a compact set $K \subset A$ such that $\mu(A \setminus K) < \varepsilon$ and that the restricted function $f|_K$ is a continuous function.

THE MORAL: A measurable function can be measure theoretically approximated by a continuous function.

Remarks 2.53:

(1) The assumption $\mu(A) < \infty$ can be removed if the compact set in the claim is replaced with a closed set.

(2) Lusin's theorem gives a characterization for measurable functions.

Reason. Assume that for every $i = 1, 2, \ldots$, there is a compact set $K_i \subset A$ such that $\mu(A \setminus K_i) < \frac{1}{i}$ and $f|_{K_i}$ is continuous. Let $B = \bigcup_{i=1}^{\infty} K_i$ and $N = A \setminus B$. Then

$$\begin{aligned} 0 \leqslant \mu(N) = \mu(A \setminus B) &= \mu\left(A \setminus \bigcup_{i=1}^{\infty} K_i\right) \\ &= \mu\left(\bigcap_{i=1}^{\infty} (A \setminus K_i)\right) \leqslant \mu(A \setminus K_i) < \frac{1}{i} \end{aligned}$$

for every $i = 1, 2, \ldots$. Thus $\mu(N) = 0$. Then

$$\{x \in A : f(x) > a\} = \{x \in B : f(x) > a\} \cup \{x \in A \setminus B : f(x) > a\}$$

for every $a \in \mathbb{R}$. The set $\{x \in B : f(x) > a\}$ is μ-measurable, since f is continuous in B and $\{x \in A \setminus B : f(x) > a\}$ is μ-measurable, since it is a set of measure zero. This implies that f is μ-measurable in A. ■

Proof. For every $i = 1, 2, \ldots$, let $B_{i,j}$, $j = 1, 2, \ldots$, be disjoint Borel sets such that

$$\bigcup_{j=1}^{\infty} B_{i,j} = \mathbb{R} \quad \text{and} \quad \operatorname{diam}(B_{i,j}) < \frac{1}{i}.$$

Denote $A_{i,j} = A \cap f^{-1}(B_{i,j})$. By Lemma 2.5, the set $A_{i,j}$ is μ-measurable. Moreover

$$A = A \cap f^{-1}(\mathbb{R}) = A \cap f^{-1}\left(\bigcup_{j=1}^{\infty} B_{i,j}\right) = \bigcup_{j=1}^{\infty} A \cap f^{-1}(B_{i,j}) = \bigcup_{j=1}^{\infty} A_{i,j}, \quad i = 1,2,\dots$$

Since $\mu(A) < \infty$, $\nu = \mu\lfloor A$ is a Radon measure by Lemma 1.42. By Corollary 1.46, there exists a compact set $K_{i,j} \subset A_{i,j}$ such that

$$\nu(A_{i,j} \setminus K_{i,j}) < \frac{\varepsilon}{2^{i+j}}$$

Then

$$\begin{aligned}\mu\left(A \setminus \bigcup_{j=1}^{\infty} K_{i,j}\right) &= \nu\left(A \setminus \bigcup_{j=1}^{\infty} K_{i,j}\right) = \nu\left(\bigcup_{j=1}^{\infty} A_{i,j} \setminus \bigcup_{j=1}^{\infty} K_{i,j}\right) \\ &\leqslant \nu\left(\bigcup_{j=1}^{\infty} (A_{i,j} \setminus K_{i,j})\right) \leqslant \sum_{j=1}^{\infty} \nu(A_{i,j} \setminus K_{i,j}) < \frac{\varepsilon}{2^i}.\end{aligned}$$

Since $\mu(A) < \infty$, Theorem 1.20 implies

$$\lim_{k\to\infty} \mu\left(A \setminus \bigcup_{j=1}^{k} K_{i,j}\right) = \mu\left(A \setminus \bigcup_{j=1}^{\infty} K_{i,j}\right) < \frac{\varepsilon}{2^i}$$

and thus there exists an index k_i such that

$$\mu\left(A \setminus \bigcup_{j=1}^{k_i} K_{i,j}\right) < \frac{\varepsilon}{2^i}$$

The set $K_i = \bigcup_{j=1}^{k_i} K_{i,j}$ is compact. For every i,j, we choose $\alpha_{i,j} \in B_{i,j}$. Then we define a function $g_i : K_i \to \mathbb{R}$ by

$$g_i(x) = \alpha_{i,j}, \quad \text{when} \quad x \in K_{i,j}, \quad j = 1, \dots k_i.$$

Since $K_{i,1}, \dots, K_{i,k_i}$ are pairwise disjoint compact sets,

$$\operatorname{dist}(K_{i,j}, K_{i,l}) > 0 \quad \text{when} \quad j \neq l.$$

This implies that g_i is continuous in K_i and

$$|f(x) - g_i(x)| < \frac{1}{i} \quad \text{for every} \quad x \in K_i,$$

since $f(K_{i,j}) \subset f(A_{i,j}) \subset B_{i,j}$ and $\operatorname{diam}(B_{i,j}) < \frac{1}{i}$. The set $K = \bigcap_{i=1}^{\infty} K_i$ is compact and

$$\begin{aligned}\mu(A \setminus K) &= \mu\left(A \setminus \bigcap_{i=1}^{\infty} K_i\right) = \mu\left(\bigcup_{i=1}^{\infty} (A \setminus K_i)\right) \\ &\leqslant \sum_{i=1}^{\infty} \mu(A \setminus K_i) < \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} = \varepsilon.\end{aligned}$$

Since

$$|f(x) - g_i(x)| < \frac{1}{i} \quad \text{for every} \quad x \in K, \; i = 1,2,\dots,$$

we see that $g_i \to f$ uniformly in K. The function f is continuous in K as a uniform limit of continuous functions. □

WARNING: Note carefully, that $f|_K$ denotes the restriction of f to K. Theorem 2.52 states that f is continuous viewed as a function defined only on the set K. This does not immediately imply that f defined as a function on A is continuous at the points in K.

Reason. $f:[0,1]\to\mathbb{R}$, $f(x)=\chi_{\mathbb{Q}}(x)$ is discontinuous at every point of $[0,1]$. However, $f|_{[0,1]\cap\mathbb{Q}}=1$ and $f_{[0,1]\setminus\mathbb{Q}}=0$ are continuous functions. It is an exercise to construct the compact set in Lusin's theorem for this function. ■

Keeping this example in mind, we are now ready to prove a stronger result.

Corollary 2.54. Let μ be a Borel regular outer measure on $\mathbb{R}^n$, $A\subset\mathbb{R}^n$ a μ-measurable set such that $\mu(A)<\infty$ and $f:\mathbb{R}^n\to\mathbb{R}$ be a μ-measurable function such that $|f|<\infty$ μ-almost everywhere. Then for every $\varepsilon>0$ there exists a continuous function $\overline{f}:\mathbb{R}^n\to\mathbb{R}$ such that

$$\mu(\{x\in A:\overline{f}(x)\neq f(x)\})<\varepsilon.$$

WARNING: The corollary does not imply that there is a continuous function $\overline{f}:\mathbb{R}^n\to\mathbb{R}$ such that $\overline{f}(x)=f(x)$ μ-almost everywhere, see Example 3.34.

Proof. Let $\varepsilon>0$. By Lusin's theorem 2.52, there exists a compact set $K\subset A$ such that $\mu(A\setminus K)<\varepsilon$ and $f|_K$ is continuous. By Tieze's extension theorem there exists a continuous function $\overline{f}:\mathbb{R}^n\to\mathbb{R}$ such that $\overline{f}(x)=f(x)$ for every $x\in K$. We refer to [2] for Tieze's theorem. Then

$$\mu(\{x\in A:\overline{f}(x)\neq f(x)\})\leqslant\mu(A\setminus K)<\varepsilon,$$

which implies the claim. □

Remarks 2.55:

(1) Tieze's extension theorem holds in metric spaces. Let F be a closed subset of a metric space X and suppose that $f:F\to\mathbb{R}$ is a continuous function. Then f can be extended to a continuous function $\overline{f}:X\to\mathbb{R}$ defined everywhere on X. Moreover, if $|f(x)|\leqslant M$ for every $x\in F$, then if $|\overline{f}(x)|\leqslant M$ for every $x\in X$. See [2].

(2) It is essential in Tieze's extension theorem that the set F is closed.

Reason. The function $f:(0,1]\to\mathbb{R}$, $f(x)=\sin\frac{1}{x}$ is a continuous function on $(0,1]$, but it cannot be extended to a continuous function to $[0,1]$. ■

Example 2.56. Let $A\subset\mathbb{R}^n$ be a Lebesgue measurable set with $m(A)<\infty$ and $f=\chi_A$. By Theorem 1.59, for every $\varepsilon>0$, there exists a compact set $K\subset A$ and an open set $G\supset A$ such that $m(A\setminus K)<\varepsilon/2$ and $m(G\setminus A)<\varepsilon/2$. As in Remark 1.29, define

$$\overline{f}(x)=\frac{\operatorname{dist}(x,\mathbb{R}^n\setminus G)}{\operatorname{dist}(x,\mathbb{R}^n\setminus G)+\operatorname{dist}(x,K)}.$$

Then $\overline{f}$ is a continuous function in $\mathbb{R}^n$ and

$$m(\{x \in \mathbb{R}^n : \overline{f}(x) \neq f(x)\}) \leqslant m(G \setminus K) = m(G \setminus A) + m(A \setminus K) < \varepsilon.$$

In this special case, the function in the previous corollary can be constructed explicitely.

The integral is first defined for nonnegative simple functions, then for nonnegative measurable functions and finally for signed functions. The integral has all basic properties one might expect and it behaves well with respect to limits, as the monotone convergence theorem, Fatou's lemma and the dominated convergence theorem show.

3 Integration

3.1 Integral of a nonnegative simple function

Let A be a μ-measurable set. It is natural to define the integral of the characteristic function of A as

$$\int_X \chi_A \, d\mu = \mu(A).$$

The same approach can be applied for simple functions. Recall that a function $f : X \to \mathbb{R}$ is simple, if its range is a finite set $\{a_1, \dots, a_n\}$, $n \in \mathbb{N}$, and the preimages

$$f^{-1}(\{a_i\}) = \{x \in X : f(x) = a_i\}$$

are μ-measurable sets, see Definition 2.31. A simple function is a linear combination of finitely many characteristic functions of μ-measurable sets, since it can be written as a finite sum

$$f = \sum_{i=1}^{n} a_i \chi_{A_i}, \quad n \in \mathbb{N},$$

where $A_i = f^{-1}(\{a_i\})$. Remark 2.2 (3) and Theorem 2.13 imply that a simple function is μ-measurable. This is called the canonical representation of a simple function. Observe that the sets A_i are disjoint and thus for each $x \in X$ there is only one nonzero term in the sum above.

Definition 3.1. Let μ be a measure on X and let $f = \sum_{i=1}^{n} a_i \chi_{A_i}$ be the canonical representation of a nonnegative simple function. Then

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

If for some i we have $a_i = 0$ and $\mu(A_i) = \infty$, we define $a_i \mu(A_i) = 0$.

THE MORAL: The definition of the integral of a simple functions is based on a subdivision of the range instead of the domain, as in the case of step functions. This is compatible with the definition of a measurable function.

Example 3.2. The function $f:\mathbb{R}\to\mathbb{R}$, $f(x)=\chi_{\mathbb{Q}}(x)$ is simple with respect to the one dimensional Lebesgue measure and $\int_{\mathbb{R}} f(x)\,dx=0$.

Remarks 3.3:

(1) For a nonnegative simple function f, we have $0\leqslant\int_X f\,d\mu\leqslant\infty$.

(2) If f is a simple function and A is μ-measurable subset of X, then $f\chi_A$ is a simple function.

(3) (Compatibility with the measure) If A is μ-measurable subset of X, then $\int_X \chi_A\,d\mu=\mu(A)$.

(4) The representation of a simple function is not in general unique in the sense that there may be several ways to write the function as a finite linear combination of characteristic functions of pairwise disjoint measurable sets. For example, $\chi_X=\chi_{X\setminus A}+\chi_A$ for every μ-measurable set $A\subset X$. However, the definition of the integral of a nonnegative simple function is independent of the representation of the function.

Reason. Let $f=\sum_{i=1}^n a_i\chi_{A_i}$ is the canonical representation of a nonnegative simple function f and let $f=\sum_{j=1}^m b_j\chi_{B_j}$ be another representation, where b_j are nonnegative real numbers and B_j pairwise disjoint μ-measurable subsets of X with $\bigcup_{j=1}^m B_j=X$. The additivity of μ on pairwise disjoint μ-measurable sets and the fact that $a_i=b_j$ if $A_i\cap B_j\neq\emptyset$ imply

$$\begin{aligned}\sum_{i=1}^n a_i\mu(A_i)&=\sum_{i=1}^n\sum_{j=1}^m a_i\mu(A_i\cap B_j)\\&=\sum_{i=1}^n\sum_{j=1}^m b_j\mu(A_i\cap B_j)\\&=\sum_{j=1}^m\sum_{i=1}^n b_j\mu(A_i\cap B_j)=\sum_{j=1}^m b_j\mu(B_j).\end{aligned}$$ ■

(5) If A is a μ-measurable subset of X, then we define

$$\int_A f\,d\mu=\int_X f\chi_A\,d\mu.$$

If $f=\sum_{i=1}^n a_i\chi_{A_i}$ is the canonical representation of a nonnegative simple function

$$\int_A f\,d\mu=\sum_{i=1}^n a_i\mu(A_i\cap A).$$

Observe that the sum on the right-hand side is not necessarily the canonical form of $f\chi_A$. The integral of a nonnegative simple function is independent of the representation.

(6) If f and g are nonnegative simple functions such that $f = g$ μ-almost everywhere, then $\int_X f\,d\mu = \int_X g\,d\mu$. Note that the converse is not true.

(7) $\int_X f\,d\mu = 0$ if and only if $f = 0$ μ-almost everywhere.

Lemma 3.4. Assume that f and g are nonnegative simple functions on X.

(1) (Monotonicity in sets) If A and B are μ-measurable sets with $A \subset B$, then $\int_A f\,d\mu \leqslant \int_B f\,d\mu$.

(2) (Homogeneity) $\int_X af\,d\mu = a\int_X f\,d\mu$, $a \geqslant 0$.

(3) (Linearity) $\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$.

(4) (Monotonicity in functions) $f \leqslant g$ implies $\int_X f\,d\mu \leqslant \int_X g\,d\mu$.

Proof. Claims (1) and (2) are clear. To prove (3), let

$$\int_X f\,d\mu = \sum_{i=1}^{n} a_i\mu(A_i) \quad \text{and} \quad \int_X g\,d\mu = \sum_{j=1}^{m} b_j\mu(B_j)$$

be the canonical representations of f and g. We have $X = \bigcup_{i=1}^{n} A_i = \bigcup_{j=1}^{m} B_j$. Then $f+g$ is a nonnegative simple function. The sets

$$C_{i,j} = A_i \cap B_j, \quad i = 1,\dots,n,\, j = 1,\dots,m,$$

are pairwise disjoint and $X = \bigcup_{i=1}^{n}\bigcup_{j=1}^{m} C_{i,j}$ and each of the functions f and g are constant on each set $C_{i,j}$. Thus

$$\begin{aligned}
\int_X (f+g)\,d\mu &= \sum_{i=1}^{n}\sum_{j=1}^{m}(a_i+b_j)\mu(C_{i,j})\\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\mu(A_i\cap B_j) + \sum_{i=1}^{n}\sum_{j=1}^{m} b_j\mu(A_i\cap B_j)\\
&= \sum_{i=1}^{n} a_i\mu(A_i) + \sum_{j=1}^{m} b_j\mu(B_j)\\
&= \int_X f\,d\mu + \int_X g\,d\mu.
\end{aligned}$$

To prove (4) we note that on the sets $C_{i,j} = A_i \cap B_j$ we have $f = a_i \leqslant b_j = g$ and thus

$$\int_X f\,d\mu = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\mu(C_{i,j}) \leqslant \sum_{i=1}^{n}\sum_{j=1}^{m} b_j\mu(C_{i,j}) = \int_X g\,d\mu.$$

□

Remark 3.5. Since the sum in the normal representation of a nonnegative simple function consists of finitely terms, it is clear that the integral inherits the properties of the measure. For example, if A_i, $i = 1,2,\dots$ are μ-measurable sets, then

$$\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu \leqslant \sum_{i=1}^{\infty}\int_{A_i} f\,d\mu$$

and

$$\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu = \sum_{i=1}^{\infty} \int_{A_i} f\,d\mu$$

if the sets A_i, $i = 1,2,\dots$, are pairwise disjoint. Moreover, if $A_i \supset A_{i+1}$ for every i and $\int_{A_1} f\,d\mu < \infty$, we have

$$\int_{\bigcap_{i=1}^{\infty} A_i} f\,d\mu = \lim_{i\to\infty} \int_{A_i} f\,d\mu$$

and finally if $A_i \subset A_{i+1}$ for every i, then

$$\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu = \lim_{i\to\infty} \int_{A_i} f\,d\mu.$$

3.2 Integral of a nonnegative measurable function

The integral of an arbitrary nonnegative measurable function is defined through an approximation by simple functions.

Definition 3.6. Let $f : X \to [0,\infty]$ be a nonnegative μ-measurable function. The integral of f with respect to μ is

$$\int_X f\,d\mu = \sup\left\{\int_X g\,d\mu : g \text{ is simple and } 0 \leqslant g(x) \leqslant f(x) \text{ for every } x \in X\right\}.$$

A nonnegative function is integrable, if

$$\int_X f\,d\mu < \infty.$$

T H E M O R A L : The integral is defined for all nonnegative measurable functions. Observe, that the integral may be ∞.

Remarks 3.7:

(1) As before, if A is a μ-measurable subset of X, then we define

$$\int_A f\,d\mu = \int_X \chi_A f\,d\mu.$$

Thus by taking the zero extension, we may assume that the function is defined on the whole space.

(2) The definition is consistent with the one for nonnegative simple functions.

(3) If $\mu(X) = 0$, then $\int_X f\,d\mu = 0$ for every f.

We collect the a few basic properties of the integral of a nonnegative function below.

Lemma 3.8. Let $f, g : X \to [0,\infty]$ be μ-measurable functions.

(1) (Monotonicity in sets) If A and B are μ-measurable sets with $A \subset B$, then
$$\int_A f\,d\mu \leqslant \int_B f\,d\mu.$$

(2) (Homogeneity) $\int_X af\,d\mu = a\int_X f\,d\mu$, $a \geqslant 0$.

(3) (Linearity) $\int_X (f+g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu$.

(4) (Monotonicity in functions) $f \leqslant g$ implies $\int_X f\,d\mu \leqslant \int_X g\,d\mu$.

(5) (Tchebyshev's inequality)
$$\mu(\{x \in X : f(x) > a\}) \leqslant \frac{1}{a}\int_X f\,d\mu$$
for every $a > 0$.

WARNING: Some of the claims do not necessarily hold true for a sign changing function. However, we may consider the absolute value of a function instead. We shall return to this later.

Proof. $\boxed{(1)}$ Follows immediately from the corresponding property for nonnegative simple functions.

$\boxed{(2)}$ If $a = 0$, then
$$\int_X (0f)\,d\mu = \int_X 0\,d\mu = 0 = 0\int_X f\,d\mu.$$
Let then $a > 0$. If g is simple and $0 \leqslant g \leqslant f$, then ag is a nonnegative simple function with $ag \leqslant af$. It follows that
$$a\int_X g\,d\mu = \int_X ag\,d\mu \leqslant \int_X af\,d\mu.$$
Taking the supremum over all such functions g implies
$$a\int_X f\,d\mu \leqslant \int_X af\,d\mu.$$
Applying this inequality gives
$$\int_X af\,d\mu = a\left(\frac{1}{a}\int_X af\,d\mu\right) \leqslant a\int_X \frac{1}{a}(af)\,d\mu = a\int_X f\,d\mu.$$

$\boxed{(3)}$ Exercise, see also the remark after the monotone convergence theorem.

$\boxed{(4)}$ Let h be a simple function with $0 \leqslant h(x) \leqslant f(x)$ for every $x \in X$. Then $0 \leqslant h(x) \leqslant g(x)$ for every $x \in X$ and thus $\int_X h\,d\mu \leqslant \int_X g\,d\mu$. By taking supremum over all such functions h we have $\int_X f\,d\mu \leqslant \int_X g\,d\mu$.

$\boxed{(5)}$ Since $f \geqslant a\chi_{\{x\in X: f(x)>a\}}$, we have
$$a\mu(\{x \in X : f(x) > a\}) = \int_X a\chi_{\{x\in X: f(x)>a\}}\,d\mu \leqslant \int_X f\,d\mu.$$
□

Lemma 3.9. Let $f : X \to [0,\infty]$ be a μ-measurable function.

(1) (Vanishing) $\int_X f\,d\mu = 0$ if and only if $f = 0$ μ-almost everywhere.

(2) (Finiteness) $\int_X f\,d\mu < \infty$ implies $f < \infty$ μ-almost everywhere.

WARNING: The claim (1) is not necessarily true for a sign changing function. The converse of claim (2) is not true: $f < \infty$ μ-almost everywhere does not imply that $\int_X f\,d\mu < \infty$.

Proof. $\boxed{(1)}$ $\boxed{\Longrightarrow}$ Let

$$A_i = \left\{x \in X : f(x) > \frac{1}{i}\right\}, \quad i = 1,2,\dots$$

By Tchebyshev's inequality

$$0 \leqslant \mu(A_i) \leqslant i \int_X f\,d\mu = 0$$

which implies that $\mu(A_i) = 0$ for every $i = 1,2,\dots$. Thus

$$\mu(\{x \in X : f(x) > 0\}) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

$\boxed{\Longleftarrow}$ Since $\mu(\{x \in X : f(x) > 0\}) = 0$, we have

$$0 = \int_{\{x\in X: f(x)>0\}} \infty\,d\mu = \int_X \infty\chi_{\{x\in X: f(x)>0\}}\,d\mu \geqslant \int_X f\,d\mu \geqslant 0.$$

Thus $\int_X f\,d\mu = 0$. Another way to prove this claim is to use the definition of integral directly (exercise).

$\boxed{(2)}$ By Tchebyshev's inequality

$$\mu(\{x \in X : f(x) = \infty\}) \leqslant \mu(\{x \in X : f(x) > i\}) \leqslant \frac{1}{i}\int_X f\,d\mu \to 0$$

as $i \to \infty$ because $\int_X f\,d\mu < \infty$. □

Lemma 3.10. Let $f, g : X \to [0,\infty]$ be μ-measurable functions. If $f = g$ μ-almost everywhere then

$$\int_X f\,d\mu = \int_X g\,d\mu.$$

THE MORAL: A redefinition of a function on a set of measure zero does not affect the integral.

Proof. Let $N = \{x \in X : f(x) \neq g(x)\}$. Then $\mu(N) = 0$ and thus

$$\int_N f\,d\mu = 0 = \int_N g\,d\mu.$$

It follows that

$$\begin{aligned}\int_X f\,d\mu &= \int_{X\setminus N} f\,d\mu + \int_N f\,d\mu = \int_{X\setminus N} f\,d\mu \\ &= \int_{X\setminus N} g\,d\mu = \int_{X\setminus N} g\,d\mu + \int_N g\,d\mu = \int_X g\,d\mu.\end{aligned}$$

□

3.3 Monotone convergence theorem

Assume that $f_i : X \to [0,\infty]$, $i = 1,2,\dots$, are μ-measurable functions and that $f_i \to f$ either everywhere or μ-almost everywhere as $i \to \infty$. Then f is a μ-measurable function, see Section 2.6. Next we discuss the question whether

$$\int_X \lim_{i\to\infty} f_i \, d\mu = \lim_{i\to\infty} \int_X f_i \, d\mu.$$

In other words, is it possible to switch the order of limit and integral? We begin with moving bump examples that we have already seen before.

Examples 3.11:

(1) (Escape to horizontal infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \chi_{[i,i+1]}(x), \quad i = 1,2,\dots$$

Then $f_i(x) = 0$ as $i \to \infty$ for every $x \in \mathbb{R}$, but

$$\int_{\mathbb{R}} \lim_{i\to\infty} f_i \, dm = \int_{\mathbb{R}} 0 \, dm = 0 < 1 = m([i,i+1]) = \lim_{i\to\infty} \int_{\mathbb{R}} f_i \, dm.$$

(2) (Escape to width infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \frac{1}{i} \chi_{[0,i]}(x), \quad i = 1,2,\dots$$

Then $f_i \to 0$ uniformly in $\mathbb{R}$, but

$$\int_{\mathbb{R}} \lim_{i\to\infty} f_i \, dm = \int_{[0,\infty)} 0 \, dm = 0 < 1 = \frac{1}{i} m([0,i]) = \lim_{i\to\infty} \int_{\mathbb{R}} f_i \, dm.$$

(3) (Escape to vertical infinity) Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = i \chi_{[\frac{1}{i},\frac{2}{i}]}(x), \quad i = 1,2,\dots$$

Then $\lim_{i\to\infty} f_i(x) = 0$ for every $x \in \mathbb{R}$, but

$$\int_{\mathbb{R}} \lim_{i\to\infty} f_i \, dm = \int_{\mathbb{R}} 0 \, dm = 0 < 1 = i m\left(\left[\frac{1}{i}, \frac{2}{i}\right]\right) = \lim_{i\to\infty} \int_{\mathbb{R}} f_i \, dm.$$

Observe that the sequence is not increasing.

(4) Let $f_i : \mathbb{R} \to \mathbb{R}$, $f_i(x) = \frac{1}{i}$, $i = 1,2,\dots$ Then $\lim_{i\to\infty} f_i(x) = 0$ for every $x \in \mathbb{R}$, but

$$\int_{\mathbb{R}} \lim_{i\to\infty} f_i \, dm = 0 < \infty = \lim_{i\to\infty} \int_{\mathbb{R}} f_i \, dm.$$

This example shows that the following monotone convergence theorem does not hold for decreasing sequences of functions.

The next convergence result will be very useful.

Theorem 3.12 (Monotone convergence theorem). If $f_i : X \to [0,\infty]$ are μ-measurable functions such that $f_i \leqslant f_{i+1}$, $i = 1,2,\dots$, then

$$\int_X \lim_{i\to\infty} f_i \, d\mu = \lim_{i\to\infty} \int_X f_i \, d\mu$$

THE MORAL: The order of taking limit and integral can be switched for an increasing sequence of nonnegative measurable functions.

Remarks 3.13:

(1) It is enough to assume that $f_i \leqslant f_{i+1}$ almost everywhere.

(2) The limits may be infinite.

(3) In the special case when $f_i = \chi_{A_i}$, where A_i is μ-measurable and $A_i \subset A_{i+1}$, the monotone convergence theorem reduces to the upwards monotone convergence result for measures, see Theorem 1.20.

Proof. Let $f = \lim_{i\to\infty} f_i$. By monotonicity

$$\int_X f_i \, d\mu \leqslant \int_X f_{i+1} \, d\mu \leqslant \int_X f \, d\mu$$

for every $i = 1,2,\dots$ This implies that the limit exists and

$$\lim_{i\to\infty} \int_X f_i \, d\mu \leqslant \int_X f \, d\mu.$$

To prove the reverse inequality, let g be a nonnegative simple function with $g \leqslant f$. Let $0 < t < 1$ and

$$A_i = \{x \in X : f_i(x) \geqslant tg(x)\}, \quad i = 1,2,\dots$$

By Lemma 2.10 and Remark 2.10, the sets A_i are μ-measurable and $A_i \subset A_{i+1}$, $i = 1,2\dots.$

Claim: $\bigcup_{i=1}^\infty A_i = X$.

Reason. $\boxed{\subset}$ Since $A_i \subset X$, $i = 1,2,\dots$, we have $\bigcup_{i=1}^\infty A_i \subset X$.

$\boxed{\supset}$ For every $x \in X$, either $f(x) \leqslant tg(x)$ or $f(x) > tg(x)$. If $f(x) \leqslant tg(x)$, then $f(x) \leqslant tg(x) \leqslant tf(x)$ and, since $0 < t < 1$, we have $f(x) = 0$. In this case $x \in A_i$ for every $i = 1,2,\dots.$ In the other hand, if $f(x) > tg(x)$, then $f(x) = \lim_{i\to\infty} f_i(x) > tg(x)$. Thus there exists i such that $f_i(x) > tg(x)$ and consequently $x \in A_i$. This shows that $x \in \bigcup_{i=1}^\infty A_i$ for every $x \in X$. ■

Thus

$$\int_X f_i \, d\mu \geqslant \int_{A_i} f_i \, d\mu \geqslant \int_{A_i} tg \, d\mu = t\int_{A_i} g \, d\mu \to t\int_X g \, d\mu$$

as $i \to \infty$. Here we used the fact that $\bigcup_{i=1}^\infty A_i = X$ and the measure properties of the integral of nonnegative simple functions, see Remark 3.5. This implies

$$\lim_{i\to\infty} \int_X f_i \, d\mu \geqslant t\int_X g \, d\mu.$$

By taking the supremum over all nonnegative simple functions g we have

$$\lim_{i\to\infty}\int_X f_i\,d\mu \geqslant t\int_X f\,d\mu$$

and the claim follows by passing $t \to 1$. □

Remarks 3.14:

(1) By Theorem 2.32 for every nonnegative μ-measurable function f there is an increasing sequence f_i, $i = 1,2,\dots$, of simple functions such that

$$f(x) = \lim_{i\to\infty} f_i(x)$$

for every $x \in X$. By the monotone convergence theorem we have

$$\int_X f\,d\mu = \lim_{i\to\infty}\int_X f_i\,d\mu.$$

Conversely, if f_i, $i = 1,2,\dots$, are nonnegative simple functions such that $f_i \leqslant f_{i+1}$ and $f = \lim_{i\to\infty} f_i$, then

$$\int_X f\,d\mu = \lim_{i\to\infty}\int_X f_i\,d\mu.$$

Moreover, this limit is independent of the approximating sequence.

(2) Let $f,g : X \to [0,\infty]$ be μ-measurable functions. Let f_i, $i = 1,2,\dots$, be an increasing sequence of nonnegative simple functions such that $f = \lim_{i\to\infty} f_i$ and let g_i, $i = 1,2,\dots$, be an increasing sequence of nonnegative simple functions such that $g = \lim_{i\to\infty} g_i$. Then

$$f(x)+g(x) = \lim_{i\to\infty}(f_i(x)+g_i(x))$$

and the monotone convergence theorem implies

$$\begin{aligned}\int_X (f+g)\,d\mu &= \lim_{i\to\infty}\int_X (f_i+g_i)\,d\mu\\ &= \lim_{i\to\infty}\int_X f_i\,d\mu + \lim_{i\to\infty}\int_X g_i\,d\mu\\ &= \int_X f\,d\mu + \int_X g\,d\mu.\end{aligned}$$

This shows the approximation by simple functions can be used to prove properties of the integral, compare to Lemma 3.8.

Example 3.15. The monotone convergence theorem can be used to compute limits of certain nonnegative integrals.

(1) Consider

$$\lim_{i\to\infty}\int_0^1 \frac{i}{1+i\sqrt{x}}\,dx.$$

Let

$$f_i(x) = \frac{i}{1+i\sqrt{x}}$$

for every $0 \leqslant x \leqslant 1$ and $i = 1,2,\dots$. Note that $0 \leqslant f_i(x) \leqslant f_{i+1}(x)$ for every $0 \leqslant x \leqslant 1$ and $i = 1,2,\dots$. The monotone convergence theorem implies

$$\begin{aligned}\lim_{i\to\infty}\int_0^1 \frac{i}{1+i\sqrt{x}}\,dx &= \lim_{i\to\infty}\int_0^1 f_i(x)\,dx = \int_0^1 \lim_{i\to\infty} f_i(x)\,dx\\ &= \int_1^2 \lim_{i\to\infty}\frac{i}{1+i\sqrt{x}}\,dx = \int_1^2 \lim_{i\to\infty}\frac{1}{\frac{1}{i}+\sqrt{x}}\,dx\\ &= \int_1^2 \frac{1}{\sqrt{x}}\,dx = 2.\end{aligned}$$

(2) Consider

$$\lim_{x\to 0+}\int_0^\infty \frac{e^{-xt}}{1+t^2}\,dt.$$

Note that $\lim_{x\to 0+} f(x)$ exists if and only if $\lim\limits_{i\to\infty} f(x_i)$ exists for all decreasing sequences (x_i) with $x_i \searrow 0$ and is independent of the sequence. Let (x_i) be such a sequence and let

$$f_i(t) = \frac{e^{-x_i t}}{1+t^2}$$

for every $t \geqslant 0$ and $i = 1,2,\dots$. Since (x_i) is a decreasing sequence, we have $0 \leqslant f_i(t) \leqslant f_{i+1}(t)$ for every $t \geqslant 0$ and $i = 1,2,\dots$. We can thus apply the monotone convergence theorem, use the fact that $\lim\limits_{i\to\infty} x_i = 0$ and continuity of elementary functions to obtain

$$\begin{aligned}\lim_{i\to\infty}\int_0^\infty \frac{e^{-x_i t}}{1+t^2}\,dt &= \lim_{i\to\infty}\int_0^\infty f_i(t)\,dt = \int_0^\infty \lim_{i\to\infty} f_i(t)\,dt\\ &= \int_0^\infty \lim_{i\to\infty}\frac{e^{-x_i t}}{1+t^2}\,dt = \int_0^\infty \frac{e^0}{1+t^2}\,dt\\ &= \int_0^\infty \frac{1}{1+t^2}\,dt = \arctan t\Big|_{t=0}^{\infty} = \frac{\pi}{2}.\end{aligned}$$

In the examples above we assumed that the familiar rules for computing integrals hold. This is a consequence of the fact that Lebesgue integral is equal to Riemann integral for bounded continuous functions, see Section 3.9.

Corollary 3.16. Let $f_i : X \to [0,\infty]$, $i = 1,2,\dots$, be nonnegative μ-measurable functions. Then

$$\int_X \sum_{i=1}^\infty f_i\,d\mu = \sum_{i=1}^\infty \int_X f_i\,d\mu.$$

THE MORAL: A series of nonnegative measurable functions can be integrated termwise.

Proof. Let $s_n = f_1 + \dots + f_n = \sum_{i=1}^n f_i$ be the nth partial sum and

$$f = \lim_{n\to\infty} s_n = \sum_{i=1}^\infty f_i.$$

The functions s_n, $n = 1,2,\dots$, form an increasing sequence of nonnegative μ-measurable functions. By the monotone convergence theorem

$$\int_X f\,d\mu = \lim_{n\to\infty}\int_X s_n\,d\mu = \lim_{n\to\infty}\sum_{i=1}^{n}\int_X f_i\,d\mu = \sum_{i=1}^{\infty}\int_X f_i\,d\mu. \qquad \square$$

Remark 3.17. Let f be a nonnegative μ-measurable function on X. Define

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

for any μ-measurable set A. Then ν is a measure.

Reason. It is clear that ν is nonnegative and that $\nu(\emptyset) = 0$. We show that ν is countably additive on pairwise disjoint μ-measurable sets. Let A_i, $i = 1,2,\dots$, be pairwise disjoint μ-measurable sets and let $f_i = f\chi_{A_i}$. Then by the previous corollary,

$$\begin{aligned}\sum_{i=1}^{\infty}\nu(A_i) &= \sum_{i=1}^{\infty}\int_X f_i\,d\mu = \int_X\sum_{i=1}^{\infty} f_i\,d\mu = \int_X\sum_{i=1}^{\infty} f\chi_{A_i}\,d\mu\\ &= \int_X f\sum_{i=1}^{\infty}\chi_{A_i}\,d\mu = \int_X f\chi_{\bigcup_{i=1}^{\infty}A_i}\,d\mu\\ &= \int_{\bigcup_{i=1}^{\infty}A_i} f\,d\mu = \nu\left(\bigcup_{i=1}^{\infty}A_i\right). \qquad \blacksquare\end{aligned}$$

This provides a useful method of constructing measures related to a nonnegative weight function f.

The properties of the measure give several useful results for integrals of a nonnegative function. These properties can also be proved using the monotone convergence theorem, compare to Remark 3.29.

(1) (Countable additivity) If A_i, $i = 1,2,\dots$ are pairwise disjoint μ-measurable sets, then

$$\int_{\bigcup_{i=1}^{\infty}A_i} f\,d\mu = \sum_{i=1}^{\infty}\int_{A_i} f\,d\mu.$$

(2) (Countable subadditivity) If A_i, $i = 1,2,\dots$ are μ-measurable sets, then

$$\int_{\bigcup_{i=1}^{\infty}A_i} f\,d\mu \leqslant \sum_{i=1}^{\infty}\int_{A_i} f\,d\mu.$$

(3) (Downwards monotone convergence) If A_i, $i = 1,2,\dots$, are μ-measurable, $A_i \supset A_{i+1}$ for every i and $\int_{A_1} f\,d\mu < \infty$, then

$$\int_{\bigcap_{i=1}^{\infty}A_i} f\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.$$

(4) (Upwards monotone convergence) If A_i, $i = 1,2,\dots$, are μ-measurable and $A_i \subset A_{i+1}$ for every i, then

$$\int_{\bigcup_{i=1}^{\infty}A_i} f\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.$$

3.4 Fatou's lemma

The next convergence result holds without monotonicity assumptions.

Theorem 3.18 (Fatou's lemma). If $f_i : X \to [0,\infty]$, $i = 1,2,\dots$, are μ-measurable functions, then

$$\int_X \liminf_{i\to\infty} f_i \, d\mu \leqslant \liminf_{i\to\infty} \int_X f_i \, d\mu.$$

THE MORAL: Fatou's lemma tells that the mass can be destroyed but not created in a pointwise limit of nonnegative functions as the moving bump examples show. The nonnegativity assumption is necessary. For example, consider the moving bump example with a negative sign.

Remarks 3.19:

(1) The power of Fatou's lemma is that there are no assumptions on the convergences. In particular, the limits

$$\lim_{i\to\infty} f_i \quad \text{and} \quad \lim_{i\to\infty} \int_X f_i \, d\mu$$

do not necessarily have to exist, but the corresponding limes inferiors exist for nonnegative functions.

(2) The moving bump examples show that a strict inequality may occur in Fatou's lemma.

Proof. Recall that

$$\liminf_{i\to\infty} f_i(x) = \sup_{j\geqslant 1}(\inf_{i\geqslant j} f_i(x)) = \lim_{j\to\infty}(\inf_{i\geqslant j} f_i(x)) = \lim_{j\to\infty} g_j(x),$$

where $g_j = \inf_{i\geqslant j} f_i$. The functions g_j, $j = 1,2,\dots$, form an increasing sequence of μ-measurable functions. By the monotone convergence theorem

$$\begin{aligned}\int_X \liminf_{i\to\infty} f_i \, d\mu &= \int_X \lim_{j\to\infty} g_j \, d\mu \\ &= \lim_{j\to\infty} \int_X g_j \, d\mu \\ &\leqslant \liminf_{i\to\infty} \int_X f_i \, d\mu,\end{aligned}$$

where the last inequality follows from the fact that $g_i \leqslant f_i$. □

3.5 Integral of a signed function

The integral of a signed function will be defined by considering the positive and negative parts of the function. Recall that $f^+, f^- \geqslant 0$ and $f = f^+ - f^-$.

Definition 3.20. Let $f : X \to [-\infty,\infty]$ be a μ-measurable function. If either $\int_X f^-\,d\mu < \infty$ or $\int_X f^+\,d\mu < \infty$, then the integral of f in X is defined as

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

Moreover, the function f is integrable in X, if both $\int_X f^-\,d\mu < \infty$ and $\int_X f^+\,d\mu < \infty$. In this case we denote $f \in L^1(X;\mu)$.

THE MORAL: The integral can be defined if either positive or negative parts have a finite integral. For an integrable function both have finite integrals.

Remark 3.21. (Triangle inequality) A function f is integrable if and only $|f|$ is integrable, that is, $\int_X |f|\,d\mu < \infty$. In this case,

$$\left|\int_X f\,d\mu\right| \leqslant \int_X |f|\,d\mu.$$

Reason. $\boxed{\Rightarrow}$ Assume that f is integrable in X. Since $|f| = f^+ + f^-$ and integral is linear on nonnegative functions, we have

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

It follows that $|f|$ is integrable in X.

$\boxed{\Leftarrow}$ Assume that $|f|$ is integrable in X. Since $0 \leqslant f^+ \leqslant |f|$ and $0 \leqslant f^- \leqslant |f|$ we have

$$\int_X f^+\,d\mu \leqslant \int_X |f|\,d\mu < \infty \quad \text{and} \quad \int_X f^-\,d\mu \leqslant \int_X |f|\,d\mu < \infty.$$

It follows that f is integrable in X.

Moreover,

$$\begin{aligned}\left|\int_X f\,d\mu\right| &= \left|\int_X f^+\,d\mu - \int_X f^-\,d\mu\right| \leqslant \left|\int_X f^+\,d\mu\right| + \left|\int_X f^-\,d\mu\right| \\ &= \int_X f^+\,d\mu + \int_X f^-\,d\mu = \int_X (f^+ + f^-)\,d\mu = \int_X |f|\,d\mu.\end{aligned}$$ ∎

Remarks 3.22:

(1) If $f \in L^1(X;\mu)$, then $|f| \in L^1(X;\mu)$ and by Lemma 3.9 we have $|f| < \infty$ μ-almost everywhere in X.

(2) (Majorant principle) Let $f : X \to [-\infty,\infty]$ be a μ-measurable function. If there exists a nonnegative integrable function g such that $|f| \leqslant g$ μ-almost everywhere, then f is integrable.

Reason. $\int_X |f|\,d\mu \leqslant \int_X g\,d\mu < \infty$. ∎

(3) A measure space $(X,\mathcal{M},\mu)$ with $\mu(X)=1$ is called a probability or sample space, μ a probability measure and sets belonging to $\mathcal{M}$ events. A probability measure is often denoted by P. In probability theory a measurable function is called a random variable, denoted for example by X. The integral is called the expectation or mean of X and it is written as

$$E(X)=\int X(\omega)\,dP(\omega).$$

Next we give some examples of integrals.

Examples 3.23:

(1) Let $X=\mathbb{R}^n$ and μ be the Lebesgue measure. We shall discuss properties of the Lebesgue measure in detail later.

(2) Let $X=\mathbb{N}$ and μ be the counting measure. Then all functions are μ-measurable. Observe that a function $f:\mathbb{N}\to\mathbb{R}$ is a sequence of real numbers with $x_i=f(i)$, $i=1,2,\dots$. Then

$$\int_X f\,d\mu=\sum_{i=1}^{\infty} f(i)=\sum_{i=1}^{\infty} x_i$$

and $f\in L^1(X;\mu)$ if and only if

$$\int_X |f|\,d\mu=\sum_{i=1}^{\infty}|f(i)|<\infty.$$

In other words, the integral is the sum of the series and integrability means that the series converges absolutely.

(3) Let $x_0\in X$ be a fixed point and recall that the Dirac measure at x_0 is defined as

$$\mu(A)=\begin{cases}1, & x_0\in A,\\ 0, & x_0\notin A.\end{cases}$$

Then all functions are μ-measurable. Moreover,

$$\int_X f\,d\mu=f(x_0)$$

and $f\in L^1(X;\mu)$ if and only if

$$\int_X |f|\,d\mu=|f(x_0)|<\infty.$$

Lemma 3.24. Let $f,g:X\to[-\infty,\infty]$ be integrable functions.

(1) (Homogeneity) $\int_X af\,d\mu=a\int_X f\,d\mu$, $a\in\mathbb{R}$.

(2) (Linearity) $\int_X (f+g)\,d\mu=\int_X f\,d\mu+\int_X g\,d\mu$.

(3) (Monotonicity in functions) $f\leqslant g$ implies $\int_X f\,d\mu\leqslant\int_X g\,d\mu$.

(4) (Vanishing) $\mu(X)=0$ implies $\int_X f\,d\mu=0$.

(5) (Almost everywhere equivalence) If $f = g$ μ-almost everywhere in X, then $\int_X f\,d\mu = \int_X g\,d\mu$.

WARNING: Monotonicity in sets does not necessarily hold for sign changing functions.

Remark 3.25. Since $f, g \in L^1(X;\mu)$, we have $|f| < \infty$ and $|g| < \infty$ μ-almost everywhere in X. Thus $f + g$ is defined μ-almost everywhere in X.

Proof. $\boxed{(1)}$ If $a \geqslant 0$, then $(af)^+ = af^+$ and $(af)^- = af^-$. This implies

$$\int_X (af)^+\,d\mu = a\int_X f^+\,d\mu \quad \text{and} \quad \int_X (af)^-\,d\mu = a\int_X f^-\,d\mu$$

The claim follows from this. If $a < 0$, then $(af)^+ = (-a)f^-$ and $(af)^- = (-a)f^+$ and the claim follows as above.

$\boxed{(2)}$ Let $h = f + g$. Then h is defined almost everywhere and measurable. The pointwise inequality $|h| \leqslant |f| + |g|$ implies

$$\int_X |h|\,d\mu \leqslant \int_X |f|\,d\mu + \int_X |g|\,d\mu < \infty$$

and thus h is integrable. Note that in general $h^+ \neq f^+ + g^+$, but

$$h^+ - h^- = h = f + g = f^+ - f^- + g^+ - g^-$$

implies

$$h^+ + f^- + g^- = h^- + f^+ + g^+.$$

Both sides are nonnegative integrable functions. It follows that

$$\int_X h^+\,d\mu + \int_X f^-\,d\mu + \int_X g^-\,d\mu = \int_X h^-\,d\mu + \int_X f^+\,d\mu + \int_X g^+\,d\mu$$

and since all integrals are finite we arrive at

$$\begin{aligned}\int_X h\,d\mu &= \int_X h^+\,d\mu - \int_X h^-\,d\mu \\ &= \int_X f^+\,d\mu - \int_X f^-\,d\mu + \int_X g^+\,d\mu - \int_X g^-\,d\mu \\ &= \int_X f\,d\mu + \int_X g\,d\mu.\end{aligned}$$

$\boxed{(3)}$ (1) and (2) imply that $g - f \geqslant 0$ is integrable and

$$\int_X g\,d\mu = \int_X f\,d\mu + \int_X (g-f)\,d\mu \geqslant \int_X f\,d\mu.$$

$\boxed{(4)}$ $\mu(X) = 0$ implies $\int_X f^+\,d\mu = 0$ and $\int_X f^-\,d\mu = 0$ and consequently $\int_X f\,d\mu = 0$.

$\boxed{(5)}$ If $f = g$ μ-almost everywhere in X, then $f^+ = g^+$ and $f^- = g^-$ μ-almost everywhere in X. This implies that

$$\int_X f^+\,d\mu = \int_X g^+\,d\mu \quad \text{and} \quad \int_X f^-\,d\mu = \int_X g^-\,d\mu,$$

from which the claim follows. □

3.6 Dominated convergence theorem

Now we are ready to state the principal convergence theorem in the theory of integration.

Theorem 3.26 (Dominated convergence theorem). Let $f_i : X \to [-\infty,\infty]$, $i = 1,2,\dots$, be μ-measurable functions such that $f_i \to f$ μ-almost everywhere as $i \to \infty$. If there exists an integrable function g such that $|f_i| \leqslant g$ μ-almost everywhere for every $i = 1,2,\dots$, then f is integrable and

$$\int_X f\,d\mu = \lim_{i\to\infty}\int_X f_i\,d\mu.$$

THE MORAL: The power of the theorem is that it applies to sign changing functions and there is no assumption on monotonicity. The order of taking limits and integral can be switched if there is an integrable majorant function g. Observe that the same g has to do for all functions f_i. The integrable majorant shuts down the loss of mass. Indeed, an integrable majorant does not exist in the moving bump examples.

Remark 3.27. As the moving bump examples show, see Examples 3.11, assumption about the integrable majorant is necessary.

Proof. Let

$$\begin{aligned} N =& \{x \in X : \liminf_{i\to\infty} f_i(x) \neq f(x)\} \cup \{x \in X : \limsup_{i\to\infty} f_i(x) \neq f(x)\} \\ &\cup \bigcup_{i=1}^{\infty}\{x \in X : |f_i(x)| > g(x)\}. \end{aligned}$$

Then $\mu(N) = 0$ and

$$|f(x)| = \lim_{i\to\infty}|f_i(x)| \leqslant g(x)$$

for every $x \in X \setminus N$. This implies that $f \in L^1(X \setminus N)$ and thus $f \in L^1(X)$. In the same way we see that $f_i \in L^1(X)$. Let

$$g_i(x) = \begin{cases} |f_i(x) - f(x)|, & x \in X \setminus N, \\ 0, & x \in N, \end{cases}$$

and $h = |f| + g$. Then $h \in L^1(X)$ and, for $x \in X \setminus N$,

$$\begin{aligned} h(x) - g_i(x) &= |f(x)| + g(x) - |f_i(x) - f(x)| \\ &\geqslant |f(x)| + g(x) - (|f_i(x)| + |f(x)|) \\ &= g(x) - |f_i(x)| \geqslant 0. \end{aligned}$$

Since $\lim_{i\to\infty} g_i = 0$ in X, Fatou's lemma implies

$$\begin{aligned}\int_X h\,d\mu &= \int_X \liminf_{i\to\infty}(h-g_i)\,d\mu\\ &\leqslant \liminf_{i\to\infty}\int_X (h-g_i)\,d\mu\\ &= \int_X h\,d\mu - \limsup_{i\to\infty}\int_X g_i\,d\mu.\end{aligned}$$

Since $\int_X h\,d\mu < \infty$, we have

$$\limsup_{i\to\infty}\int_X g_i\,d\mu \leqslant 0.$$

Since $g_i \geqslant 0$, we have

$$0 = \lim_{i\to\infty}\int_X g_i\,d\mu = \lim_{i\to\infty}\int_X |f_i - f|\,d\mu.$$

It follows that

$$\left|\int_X f_i\,d\mu - \int_X f\,d\mu\right| = \left|\int_X (f_i - f)\,d\mu\right| \leqslant \int_X |f_i - f|\,d\mu \to 0$$

as $i\to\infty$. □

Remarks 3.28:

(1) The proof shows that

$$\lim_{i\to\infty}\int_X |f_i - f|\,d\mu = 0.$$

This is also clear from the dominated convergence theorem, since $|f_i - f| \to 0$ μ-almost everywhere and $|f_i - f| \leqslant |f_i| + |f| \leqslant 2g$. Thus

$$\lim_{i\to\infty}\int_X |f_i - f|\,d\mu = \int_X \lim_{i\to\infty}|f_i - f|\,d\mu = 0.$$

(2) The result is interesting and useful already for the characteristic functions of measurable sets.

(3) Assume that $\mu(X) < \infty$ and $f_i : X \to [-\infty,\infty]$, $i = 1,2,\dots$, are μ-measurable functions such that $f_i \to f$ μ-almost everywhere as $i\to\infty$. If there exists $M < \infty$ such that $|f_i| \leqslant M$ μ-almost everywhere for every $i = 1,2,\dots$, then f is integrable and

$$\int_X f\,d\mu = \lim_{i\to\infty}\int_X f_i\,d\mu.$$

Reason. The constant function $g = M$ is integrable in X, since $\mu(X) < \infty$. ■

(4) Assume that $\mu(X) < \infty$ and $f_i : X \to [-\infty,\infty]$, $i = 1,2,\dots$, are integrable functions on X such that $f_i \to f$ uniformly in X as $i\to\infty$. Then f is integrable and

$$\int_X f\,d\mu = \lim_{i\to\infty}\int_X f_i\,d\mu.$$

We leave this as an exercise.

(5) We deduced the dominated convergence theorem from Fatou's lemma and Fatou's lemma from the monotone convergence theorem. This can be done in other order as well.

Remark 3.29. We have the following useful results for a function $f \in L^1(X;\mu)$. Compare these properties to the corresponding properties for nonnegative measurable functions.

(1) (Countable additivity) If A_i, $i = 1,2,\dots$, are pairwise disjoint μ-measurable sets, then

$$\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu = \sum_{i=1}^{\infty} \int_{A_i} f\,d\mu.$$

Reason. Let $s_n = \sum_{i=1}^{n} f\chi_{A_i}$, $n = 1,2,\dots$, and denote $A = \bigcup_{i=1}^{\infty} A_i$. Then $s_n \to f\chi_A$ everywhere in X as $n \to \infty$. By the triangle inequality

$$|s_n| = \left|\sum_{i=1}^{n} f\chi_{A_i}\right| \leqslant \sum_{i=1}^{n} |f|\chi_{A_i} \leqslant \sum_{i=1}^{\infty} |f|\chi_{A_i} \leqslant |f|,$$

where $f \in L^1(X;\mu)$. By the dominated convergence theorem

$$\begin{aligned}\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu &= \int_X f\chi_A\,d\mu = \int_X \lim_{n\to\infty} s_n\,d\mu \\ &= \lim_{n\to\infty}\int_X s_n\,d\mu = \lim_{n\to\infty}\sum_{i=1}^{n}\int_{A_i} f\,d\mu = \sum_{i=1}^{\infty}\int_{A_i} f\,d\mu.\end{aligned}$$

The last equality follows from the fact that the partial sums converge absolutely by the estimate above. ■

(2) (Downwards monotone convergence) If A_i is μ-measurable, $A_i \supset A_{i+1}$, $i = 1,2,\dots$, then

$$\int_{\bigcap_{i=1}^{\infty} A_i} f\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.$$

Reason. Let $f_i = f\chi_{A_i}$ and denote $A = \bigcap_{i=1}^{\infty} A_i$. Then $|f_i| \leqslant |f|$, $f \in L^1(X)$ and $f_i \to f\chi_A$ everywhere in X as $i \to \infty$. By the dominated convergence theorem

$$\begin{aligned}\int_{\bigcap_{i=1}^{\infty} A_i} f\,d\mu &= \int_X f\chi_A\,d\mu = \int_X \lim_{i\to\infty} f_i\,d\mu \\ &= \lim_{i\to\infty}\int_X f_i\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.\end{aligned}$$ ■

(3) (Upwards monotone convergence) If A_i, $i = 1,2,\dots$, are μ-measurable and $A_i \subset A_{i+1}$, $i = 1,2,\dots$, then

$$\int_{\bigcup_{i=1}^{\infty} A_i} f\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.$$

Reason. Let $f_i = f\chi_{A_i}$ and denote $A = \bigcup_{i=1}^\infty A_i$. Then $|f_i| \leqslant |f|$, $f \in L^1(X)$ and $f_i \to f\chi_A$ everywhere in X as $i \to \infty$. By the dominated convergence theorem

$$\begin{aligned}\int_{\bigcup_{i=1}^\infty A_i} f\,d\mu &= \int_X f\chi_A\,d\mu = \int_X \lim_{i\to\infty} f_i\,d\mu \\ &= \lim_{i\to\infty}\int_X f_i\,d\mu = \lim_{i\to\infty}\int_{A_i} f\,d\mu.\end{aligned}$$ ■

Example 3.30. The dominated convergence theorem can be used to compute limits of certain integrals.

(1) Consider
$$\lim_{i\to\infty}\int_0^1 ix^{-\frac{3}{2}}\sin\left(\frac{x}{i}\right)dx.$$
Let
$$f_i(x) = nx^{-\frac{3}{2}}\sin\left(\frac{x}{i}\right) = \frac{i}{x}\sin\left(\frac{x}{i}\right)x^{-\frac{1}{2}}$$
for every $0 \leqslant x \leqslant 1$ and $i = 1, 2, \dots$. Since $\lim_{x\to 0}\frac{\sin x}{x} = 1$, we have
$$\lim_{i\to\infty}\frac{i}{x}\sin\left(\frac{x}{i}\right) = 1,$$
and thus $\lim_{i\to\infty} f_i(x) = x^{-\frac{1}{2}}$ for every $0 \leqslant x \leqslant 1$. Since $|\sin t| \leqslant t$ for every $t \geqslant 0$, we have
$$\left|\frac{i}{x}\sin\left(\frac{x}{i}\right)\right| \leqslant \left|\frac{i}{x}\cdot\frac{x}{i}\right| \leqslant 1$$
and thus
$$|f_i(x)| = \left|\frac{i}{x}\sin\left(\frac{x}{i}\right)x^{-\frac{1}{2}}\right| \leqslant x^{-\frac{1}{2}}$$
for every $i = 1, 2, \dots$ and every $0 \leqslant x \leqslant 1$. Before we can use the dominated convergence theorem, we need to show that the function $g : [0,1] \to \mathbb{R}$, $g(x) = x^{-\frac{1}{2}}$ is integrable, but this is clear since
$$\int_0^1 g(x)\,dx = 2 < \infty.$$
The dominated convergence theorem implies
$$\begin{aligned}\lim_{i\to\infty}\int_0^1 ix^{-\frac{3}{2}}\sin\left(\frac{x}{i}\right)dx &= \lim_{i\to\infty}\int_0^1 f_i(x)\,dx \\ &= \int_0^1 \lim_{i\to\infty} f_i(x)\,dx = \int_0^1 g(x)\,dx = 2.\end{aligned}$$

(2) Assume $f \in L^1(\mathbb{R}^n)$ with respect to the Lebesgue measure. Then
$$\lim_{i\to\infty}\int_{\mathbb{R}^n} f(x)e^{-|x|^2/i}\,dx = \int_{\mathbb{R}^n} f(x)\,dx.$$

Reason. Since

$$\left|f(x)e^{-|x|^2/i}\right| \leqslant |f(x)|$$

for every $x \in \mathbb{R}^n$ and $i = 1,2,\dots$, the function $|f|$ will do as an integrable majorant in the dominated convergence theorem. Thus

$$\lim_{i\to\infty}\int_{\mathbb{R}^n} f(x)e^{-|x|^2/i}\,dx = \int_{\mathbb{R}^n} f(x)\underbrace{\lim_{i\to\infty} e^{-|x|^2/i}}_{=1}\,dx = \int_{\mathbb{R}^n} f(x)\,dx.$$

■

We conclude this section with two useful results, which are related to integrals depending on a parameter. Assume that μ is a measure on X and let $I \subset \mathbb{R}$ be an interval. Let $f : X \times I \to [-\infty,\infty]$, $f = f(x,t)$, be such that for every $t \in I$ the function $x \mapsto f(x,t)$ is integrable. For each $t \in I$, we consider the integral of the function over X and denote

$$F(t) = \int_X f(x,t)\,\mu(x).$$

We are interested in the regularity of F. First we discuss continuity.

Theorem 3.31 (Continuity). Assume that

(1) for every $t \in I$, the function $x \mapsto f(x,t)$ is integrable in X,

(2) the function $t \mapsto f(x,t)$ is continuous for every $x \in X$ at $t_0 \in I$ and

(3) there exists $g \in L^1(X;\mu)$ such that $|f(x,t)| \leqslant g(x)$ for every $(x,t) \in X \times I$.

Then F is continuous at t_0.

Proof. This is a direct consequence of the dominated convergence theorem, since g will do as an integrable majorant and

$$\begin{aligned}\lim_{t\to t_0} F(t) &= \lim_{t\to t_0}\int_X f(x,t)\,\mu(x)\\ &= \int_X \lim_{t\to t_0} f(x,t)\,\mu(x)\\ &= \int_X f(x,t_0)\,\mu(x) = F(t_0).\end{aligned}$$

□

THE MORAL: Under these assumptions we can take limit under the integral sign. In other words, we can switch the order of taking limit and integral.

Then we discuss differentiability.

Theorem 3.32 (Differentiability). Assume that

(1) for every $t \in I$, the function $x \mapsto f(x,t)$ is integrable in X,

(2) the function $t \mapsto f(x,t)$ is differentiable for every $x \in X$ at every point $t \in I$ and

(3) there exists $h \in L^1(X;\mu)$ such that $\left|\frac{\partial f}{\partial t}(x,t)\right| \leqslant h(x)$ for every $(x,t) \in X \times I$.

Then F is differentiable at every point $t \in I$ and

$$F'(t) = \frac{\partial}{\partial t}\left(\int_X f(x,t)\,\mu(x)\right) = \int_X \frac{\partial}{\partial t} f(x,t)\,d\mu(x).$$

T H E M O R A L : Under these assumptions we can differentiate under the integral sign. In other words, we can switch the order of taking derivative and integral.

Proof. Let $t \in I$ be fixed. For $|h|$ small consider the difference quotient

$$\frac{F(t+h)-F(t)}{h} = \int_X \frac{f(x,t+h)-f(x,t)}{h}\,d\mu(x)$$

Since f is differentiable, we have

$$\lim_{h\to 0} \frac{f(x,t+h)-f(x,t)}{h} = \frac{\partial}{\partial t} f(x,t).$$

By the mean value theorem of differential calculus

$$\left|\frac{f(x,t+h)-f(x,t)}{h}\right| = \left|\frac{\partial}{\partial t} f(x,t')\right| \leqslant h(x)$$

for some $t' \in (t,t+h)$. Thus by the dominated convergence theorem

$$\begin{aligned} F'(t) &= \lim_{h\to 0} \frac{F(t+h)-F(t)}{h} \\ &= \lim_{h\to 0} \int_X \frac{f(x,t+h)-f(x,t)}{h}\,d\mu(x) \\ &= \int_X \lim_{h\to 0} \frac{f(x,t+h)-f(x,t)}{h}\,d\mu(x) \\ &= \int_X \frac{\partial}{\partial t} f(x,t)\,d\mu(x). \end{aligned}$$

□

This kind of arguments are frequently used for the Lebesgue integral and partial derivatives in real analysis.

3.7 Lebesgue integral

Lebesgue integrable functions

Let $f : \mathbb{R}^n \to [-\infty,\infty]$ be a Lebesgue measurable function. The Lebesgue integral of f is denoted as

$$\int_{\mathbb{R}^n} f\,dm = \int_{\mathbb{R}^n} f(x)\,dm(x) = \int_{\mathbb{R}^n} f(x)\,dx = \int_{\mathbb{R}^n} f\,dx,$$

whenever the integral is defined. For a Lebesgue measurable subset A of $\mathbb{R}^n$, we define

$$\int_A f\,dx = \int_{\mathbb{R}^n} f\chi_A\,dx.$$

Example 3.33. Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = |x|^{-\alpha}$, $\alpha > 0$. The function becomes unbounded in any neighbourhood of the origin. The function f is not defined at the origin, but we may set $f(0) = 0$.

Let $A = B(0,1)$ and define $A_i = B(0,2^{-i+1}) \setminus B(0,2^{-i})$, $i = 1,2,\dots$. The sets A_i are Lebesgue measurable, pairwise disjoint and $B(0,1) = \bigcup_{i=1}^{\infty} A_i$. Thus

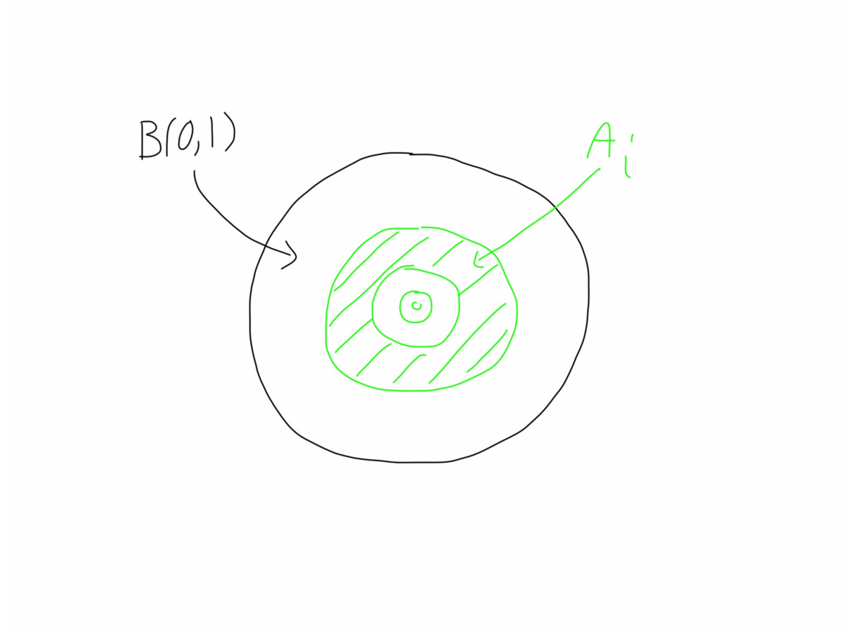

Figure 3.1: An exhaustion of $B(0,1)$ by annuli.

$$\begin{aligned}
\int_{B(0,1)} \frac{1}{|x|^\alpha}\,dx &\leqslant \sum_{i=1}^{\infty} \int_{A_i} \frac{1}{|x|^\alpha}\,dx \leqslant \sum_{i=1}^{\infty} \int_{A_i} 2^{i\alpha}\,dx = \sum_{i=1}^{\infty} 2^{i\alpha} m(A_i) \\
&\leqslant \sum_{i=1}^{\infty} 2^{i\alpha} m(B(0,2^{-i+1})) = \sum_{i=1}^{\infty} 2^{i\alpha} 2^{n(-i+1)} m(B(0,1)) \\
&= 2^n m(B(0,1)) \sum_{i=1}^{\infty} 2^{i\alpha - in} < \infty, \quad \alpha < n.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\int_{B(0,1)} \frac{1}{|x|^\alpha}\,dx &= \sum_{i=1}^{\infty} \int_{A_i} \frac{1}{|x|^\alpha}\,dx \geqslant \sum_{i=1}^{\infty} \int_{A_i} 2^{(i-1)\alpha}\,dx = \sum_{i=1}^{\infty} 2^{(i-1)\alpha} m(A_i) \\
&= (2^n - 1) 2^{-\alpha} m(B(0,1)) \sum_{i=1}^{\infty} 2^{i\alpha - in} = \infty, \quad \alpha \geqslant n.
\end{aligned}$$

The last equality follows from

$$\begin{aligned}
m(A_i) &= m(B(0,2^{-i+1})) - m(B(0,2^{-i})) \\
&= (2^{(-i+1)n} - 2^{-in}) m(B(0,1)) \\
&= 2^{-in}(2^n - 1) m(B(0,1)).
\end{aligned}$$

Thus
$$\int_{B(0,1)} \frac{1}{|x|^\alpha}\,dx < \infty \iff \alpha < n.$$
A similar reasoning with the sets $A_i = B(0,2^i) \setminus B(0,2^{i-1})$, $i = 1,2,\dots$, shows that
$$\int_{\mathbb{R}^n \setminus B(0,1)} \frac{1}{|x|^\alpha}\,dx < \infty \iff \alpha > n.$$

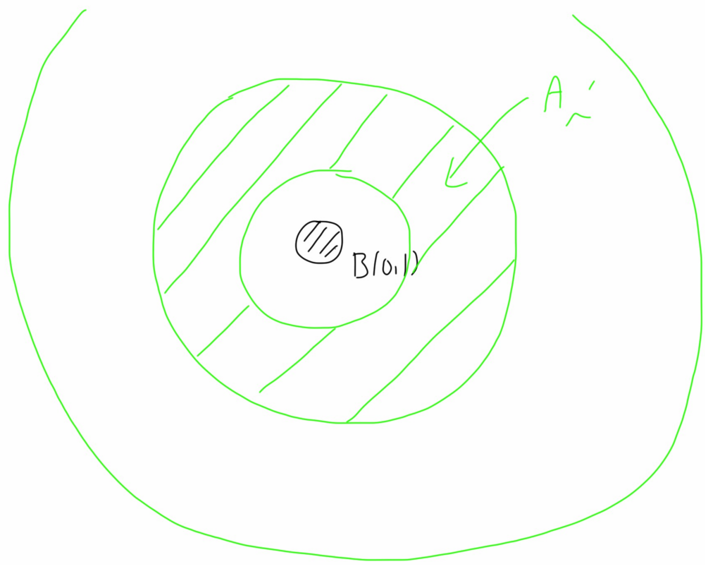

Figure 3.2: An exhaustion of $\mathbb{R}^n \setminus B(0,1)$ by annuli.

We shall show in Example 3.37 how to compute these integrals by a change of variables and spherical coordinates.

Example 3.34. Let $\phi : \mathbb{R} \to [0,\infty]$,
$$\phi(x) = \begin{cases} \frac{1}{\sqrt{x}}, & |x| < 1, \\ 0, & |x| \geqslant 1, \end{cases}$$
and define
$$f(x) = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} 2^{-i-|j|} \phi\left(x - \frac{j}{i}\right).$$

This is an infinite sum of functions with the singularities at the points $\frac{j}{i}$ with

$$\begin{aligned}\int_{\mathbb{R}} f(x)\,dx &= \sum_{i=1}^{\infty}\sum_{j=-\infty}^{\infty} 2^{-i-|j|}\int_{\mathbb{R}} \phi\left(x-\frac{j}{i}\right)dx\\ &= \sum_{i=1}^{\infty}\sum_{j=-\infty}^{\infty} 2^{-i-|j|}\int_{\mathbb{R}} \phi(x)\,dx\\ &= 4\sum_{i=1}^{\infty}\sum_{j=-\infty}^{\infty} 2^{-i-|j|} = 12 < \infty.\end{aligned}$$

Thus $f \in L^1(\mathbb{R})$. Note that f has a singularity at every rational point,

$$\lim_{x\to q} f(x) = \infty \quad \text{for every } q \in \mathbb{Q}.$$

However, since f is integrable $f(x) < \infty$ for almost every $x \in \mathbb{R}$. In other words, the series

$$f(x) = \sum_{i=1}^{\infty}\sum_{j=-\infty}^{\infty} 2^{-i-|j|}\phi\left(x-\frac{j}{i}\right)$$

converges for almost every $x \in \mathbb{R}$.

A similar example can be constructed in $\mathbb{R}^n$ (exercise). This also shows that the function f cannot be redefined on a set of measure zero so that it becomes continuous, compare to Lusin's theorem 2.52.

L^1 space

Let $f : \mathbb{R}^n \to [-\infty,\infty]$ be an integrable function in $\mathbb{R}^n$ with respect to the Lebesgue measure, denoted by $f \in L^1(\mathbb{R}^n)$. The number

$$\|f\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |f|\,dx < \infty.$$

is called the L^1-norm of f. This has the usual properties

(1) $0 \leqslant \|f\|_{L^1(\mathbb{R}^n)} < \infty$,

(2) $\|f\|_{L^1(\mathbb{R}^n)} = 0 \iff f = 0$ almost everywhere,

(3) $\|af\|_{L^1(\mathbb{R}^n)} = |a|\|f\|_{L^1(\mathbb{R}^n)}$, $a \in \mathbb{R}$, and

(4) $\|f+g\|_{L^1(\mathbb{R}^n)} \leqslant \|f\|_{L^1(\mathbb{R}^n)} + \|g\|_{L^1(\mathbb{R}^n)}$.

The last triangle inequality in L^1 follows from the pointwise triangle inequality $|f(x)+g(x)| \leqslant |f(x)|+|g(x)|$. However, there are slight problems with the vector space properties of $L^1(\mathbb{R}^n)$, since the sum function $f+g$ may be $\infty-\infty$ and is not necessarily defined at every point. However, by Lemma 3.9 (2) integrable functions are finite almost everywhere and this is not a serious problem. Moreover, $\|f\|_{L^1(\mathbb{R}^n)} = 0$ implies that $f = 0$ almost everywhere, but not necessarily everywhere, see Lemma 3.9 (1). We can overcome this problem by considering equivalence classes of functions that coincide almost everywhere.

We also recall the following useful properties which also hold for more general measures. Let $f \in L^1(\mathbb{R}^n)$. Then the following claims are true:

(1) (Finiteness) If $f \in L^1(\mathbb{R}^n)$, then $|f| < \infty$ almost everywhere in $\mathbb{R}^n$. The converse does not hold as the example above shows.

(2) (Vanishing) If $\int_{\mathbb{R}^n} |f|\,dx = 0$, then $f = 0$ almost everywhere in $\mathbb{R}^n$.

(3) (Horizontal truncation) Approximation by integrals over bounded sets

$$\int_{\mathbb{R}^n} |f|\,dx = \int_{\mathbb{R}^n} \lim_{i\to\infty} \chi_{B(0,i)}|f|\,dx = \lim_{i\to\infty} \int_{B(0,i)} |f|\,dx.$$

Here we used the monotone convergence theorem or the dominated convergence theorem if $f \in L^1(\mathbb{R}^n)$.

(4) (Vertical truncation) Approximation by integrals of bounded functions

$$\int_{\mathbb{R}^n} |f|\,dx = \int_{\mathbb{R}^n} \lim_{i\to\infty} \min\{|f|, i\}\,dx = \lim_{i\to\infty} \int_{\mathbb{R}^n} \min\{|f|, i\}\,dx.$$

Here we again used the monotone convergence theorem or the dominated convergence theorem if $f \in L^1(\mathbb{R}^n)$.

L^1 convergence

We say that $f_i \to f$ in $L^1(\mathbb{R}^n)$, if

$$\|f_i - f\|_{L^1(\mathbb{R}^n)} \to 0$$

as $i \to 0$. This is yet another mode of convergence.

Remarks 3.35:

(1) If $f_i \to f$ in $L^1(\mathbb{R}^n)$, then $f_i \to f$ in measure.

Reason. By Tchebyschev's inequality (Lemma 3.8 (5))

$$m(\{x \in \mathbb{R}^n : |f_i(x) - f(x)| > \varepsilon\}) \leqslant \frac{1}{\varepsilon}\int_{\mathbb{R}^n} |f_i(x) - f(x)|\,dx \to 0 \quad \text{as} \quad i \to \infty$$

for every $\varepsilon > 0$. ■

Theorem 2.45 implies that if $f_i \to f$ in $L^1(\mathbb{R}^n)$, then there exists a subsequence such that $f_{i_k} \to f$ μ-almost everywhere. An example of a sliding sequence of functions, see Example 2.44 (4), shows that the claim is not true for the original sequence.

(2) The Riesz-Fischer theorem states that L^1 is a Banach space, that is, every Cauchy sequence converges. We shall prove this result in the real analysis course.

Invariance properties

The invariance properties of the Lebesgue measure in Section 1.8 imply the following results:

(1) (Translation invariance)

$$\int_{\mathbb{R}^n} f(x+x_0)\,dx = \int_{\mathbb{R}^n} f(x)\,dx$$

for any $x_0 \in \mathbb{R}^n$. This means that the Lebesgue integral is invariant in translations.

Reason. We shall check this first with $f = \chi_A$, where A is Lebesgue measurable. Then $\chi_A(x+x_0) = \chi_{A-x_0}(x)$ and the claim follows from

$$\begin{aligned}\int_{\mathbb{R}^n} f(x+x_0)\,dx &= \int_{\mathbb{R}^n} \chi_A(x+x_0)\,dx = \int_{\mathbb{R}^n} \chi_{A-x_0}(x)\,dx \\ &= m(A-x_0) = m(A) \\ &= \int_{\mathbb{R}^n} \chi_A(x)\,dx = \int_{\mathbb{R}^n} f(x)\,dx.\end{aligned}$$

By linearity, the result holds for nonnegative simple functions. For nonnegative Lebesgue measurable functions the claim follows from the monotone convergence theorem by approximating with an increasing sequence of simple functions. The general case follows from this. ■

(2) (Reflection invariance)

$$\int_{\mathbb{R}^n} f(x)\,dx = \int_{\mathbb{R}^n} f(-x)\,dx.$$

(3) (Scaling property)

$$\int_{\mathbb{R}^n} f(x)\,dx = |\delta|^n \int_{\mathbb{R}^n} f(\delta x)\,dx$$

for any $\delta \neq 0$. This shows that the Lebesgue integral behaves as expected in dilations (exercise).

(4) (Linear change of variables) Let $L : \mathbb{R}^n \to \mathbb{R}^n$ be a general invertible linear mapping. Then

$$\int_{\mathbb{R}^n} f(Lx)\,dx = \frac{1}{|\det L|} \int_{\mathbb{R}^n} f(x)\,dx,$$

or equivalently,

$$\int_{\mathbb{R}^n} f(L^{-1}x)\,dx = |\det L| \int_{\mathbb{R}^n} f(x)\,dx.$$

Reason. Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set and $f = \chi_A$. Then $\chi_A \circ L = \chi_{L^{-1}(A)}$ is a Lebesgue measurable function and

$$\begin{aligned}\int_{\mathbb{R}^n} f(Lx)\,dx &= \int_{\mathbb{R}^n} \chi_A(Lx)\,dx = \int_{\mathbb{R}^n} (\chi_A \circ L)(x)\,dx = \int_{\mathbb{R}^n} \chi_{L^{-1}(A)}(x)\,dx \\ &= m(L^{-1}(A)) = \left|\det L^{-1}\right| m(A) \\ &= \frac{1}{|\det L|} \int_{\mathbb{R}^n} \chi_A(x)\,dx = \frac{1}{|\det L|} \int_{\mathbb{R}^n} f(x)\,dx.\end{aligned}$$

By taking linear combinations, we conclude the result for simple functions and the general case follows from the fact that a measurable function can be approximated by simple functions and the definition of the integral, see [7] pages 170–171 and 65–80 or or [16] pages 612–619. ■

This is a change of variables formula for linear mappings, which is compatible with the corresponding property $m(L(A)) = |\det L| m(A)$ of the Lebesgue measure, see Section 1.8.

Reason.

$$\begin{aligned} m(L(A)) &= \int_{L(A)} 1\,dx = \int_{\mathbb{R}^n} \chi_{L(A)}(x)\,dx \\ &= \int_{\mathbb{R}^n} (\chi_A \circ L^{-1})(x)\,dx = \int_{\mathbb{R}^n} \chi_A(L^{-1}x)\,dx \\ &= |\det L| \int_{\mathbb{R}^n} \chi_A(x)\,dx = |\det L| m(A). \end{aligned}$$ ■

Moreover, A is a Borel set if and only if $L(A)$ is a Borel set, since L is a homeomorphism.

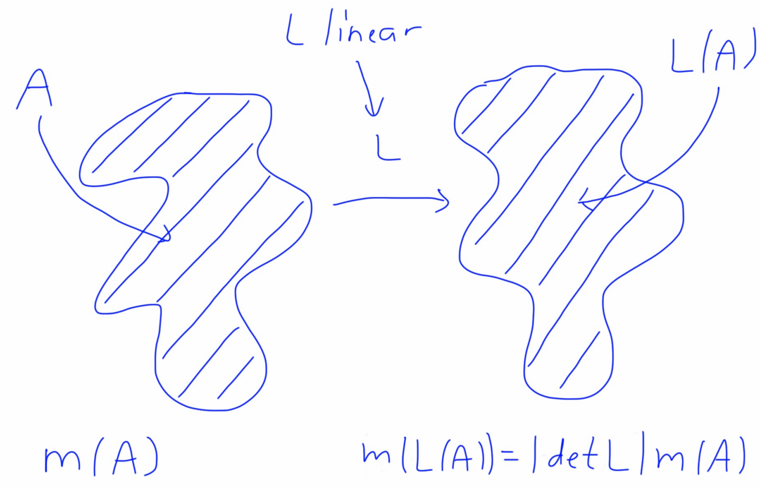

Figure 3.3: A linear change of variables.

(5) (Nonlinear change of variables) Let $U \subset \mathbb{R}^n$ be an open set and suppose that $\Phi : U \to \mathbb{R}^n$, $\Phi = (\phi_1, \dots, \phi_n)$ is a C^1 diffeomorphism. We denote by $D\Phi$ the derivative matrix with entries $D_j \phi_i$, $i, j = 1, \dots, n$. The mapping Φ is called C^1 diffeomorphism if it is injective and $D\Phi(x)$ is invertible at every $x \in U$. In this case the inverse function theorem guarantees that the

inverse map $\Phi^{-1} : \Phi(U) \to U$ is also a C^1 diffeomorphism. This means that all component functions ϕ_i, $i = 1, \dots, n$ have continuous first order partial derivatives and

$$D\Phi^{-1}(y) = (D\Phi(\Phi^{-1}(y)))^{-1}$$

for every $y \in \Phi(U)$. If f is a Lebesgue measurable function on $\Phi(U)$, then $f \circ \Phi$ is a Lebesgue measurable function on U. If f is nonnegative or integrable on $\Phi(U)$, then

$$\int_{\Phi(U)} f(y)\,dy = \int_U f(\Phi(x))|\det D\Phi(x)|\,dx.$$

Moreover, if $A \subset U$ is a Lebesgue measurable set, then $\Phi(A)$ is a Lebesgue measurable set and

$$m(\Phi(A)) = \int_A |\det D\Phi|\,dx.$$

This is a change of variables formula for differentiable mappings, see [7, p. 494–503] or [16, p. 649–660]. Formally it can be seen as the substition $y = \Phi(x)$. This means that we replace $f(y)$ by $f(\Phi(x))$, $\Phi(U)$ by U and dy by $|\det D\Phi(x)|\,dx$. Observe, that if Φ is a linear mapping, that is there exists a matrix A with $\Phi(x) = Ax$, then $D\Phi = A$, and this is compatible with the change of variables formula for linear mappings.

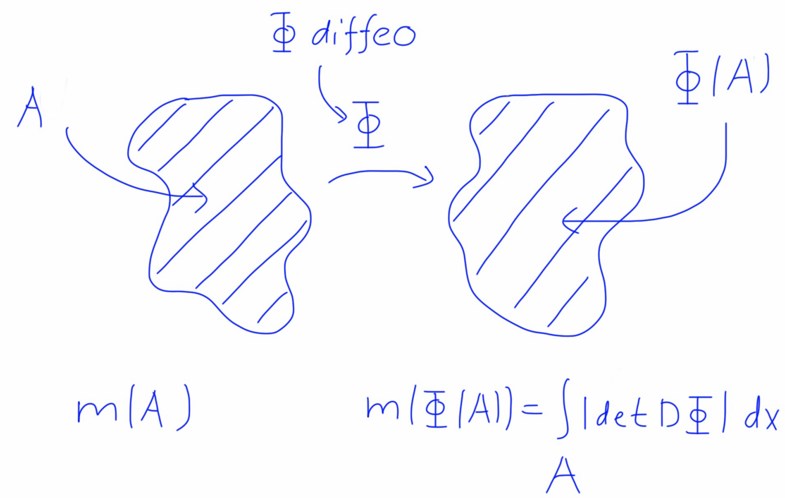

Figure 3.4: A diffeomorphic change of variables.

Example 3.36. Probably the most important nonlinear coordinate systems in $\mathbb{R}^2$ are the polar coordinates ($x_1 = r\cos\theta_1$, $x_2 = r\sin\theta_1$) and in $\mathbb{R}^3$ are the spherical coordinates ($x_1 = r\cos\theta_1$, $x_2 = r\sin\theta_1\cos\theta_2$, $x_3 = r\sin\theta_1\sin\theta_2$). Let us consider the spherical coordinates in $\mathbb{R}^n$. Let

$$U = (0,\infty)\times(0,\pi)^{n-2}\times(0,2\pi)\subset\mathbb{R}^n, \quad n\geqslant 2.$$

Denote the coordinates of a point in U by $r,\theta_1,\dots,\theta_{n-2},\theta_{n-1}$, respectively. We define $\Phi: U\to\mathbb{R}^n$ by the spherical coordinate formulas as follows. If $x=\Phi(r,\theta)$, then

$$x_i = r\sin\theta_1\cdots\sin\theta_{i-1}\cos\theta_i, \quad i=1,\dots,n,$$

where $\theta_n = 0$ so that $x_n = r\sin\theta_1\cdots\sin\theta_{n-1}$. Then ϕ is a bijection from U onto the open set $\mathbb{R}^n\setminus(\mathbb{R}^{n-1}\times[0,\infty)\times\{0\})$. The change of variables formula implies that

$$\begin{aligned}&\int_{\mathbb{R}^n} f(x)\,dx\\ &= \int_0^r\int_0^\pi\cdots\int_0^\pi\int_0^{2\pi} f(\Phi(r,\theta))r^{n-1}(\sin\theta_1)^{n-2}\dots(\sin\theta_{n-3})^2\sin\theta_{n-2}\,d\theta_{n-1}\dots d\theta_1\,dr.\end{aligned}$$

It can be shown that

$$\omega_{n-1} = \int_0^\pi(\sin\theta_1)^{n-2}\,d\theta_{n-1}\cdots\int_0^\pi(\sin\theta_{n-3})^2\,d\theta_{n-3}\int_0^\pi\sin\theta_{n-2}\,d\theta_{n-2}\int_0^{2\pi}d\theta_{n-1},$$

where

$$\omega_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

is the $(n-1)$-dimensional volume of the unit sphere $\partial B(0,1) = \{x\in\mathbb{R}^n : |x| = 1\}$. Here

$$\Gamma(a) = \int_0^\infty x^{a-1}e^{-x}\,dx, \quad 0<a<\infty,$$

is the gamma function. The gamma function has the properties $\Gamma(1) = 1$ and $\Gamma(a+1) = a\Gamma(a)$. It follows that $\Gamma(k+1) = k!$ for a nonnegative integer k.

Suppose that $f:\mathbb{R}^n\to[0,\infty]$ is radial. Thus f depends only on $|x|$ and it can be expressed as $f(|x|)$, where f is a function defined on $[0,\infty)$. Then

$$\int_{\mathbb{R}^n} f(|x|)\,dx = \omega_{n-1}\int_0^\infty f(r)r^{n-1}\,dr. \tag{3.1}$$

see [7] pages 503–504 or [16] pages 661–673.

Let us show how to use this formula to compute the volume of a ball $B(x,r) = \{y\in\mathbb{R}^n : |y-x|<r\}$, $x\in\mathbb{R}^n$ and $r>0$. Denote $\Omega_n = m(B(0,1))$. By the translation and scaling invariance and (3.1), we have

$$\begin{aligned} r^n\Omega_n &= r^n m(B(0,1)) = m(B(x,r)) = m(B(0,r))\\ &= \int_{\mathbb{R}^n}\chi_{B(0,r)}(y)\,dy = \int_{\mathbb{R}^n}\chi_{(0,r)}(|y|)\,dy\\ &= \omega_{n-1}\int_0^r\rho^{n-1}\,d\rho = \omega_{n-1}\frac{r^n}{n}.\end{aligned}$$

In particular, it follows that $\omega_{n-1} = n\Omega_n$ and

$$m(B(x,r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)}\frac{r^n}{n} = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}r^n.$$

Example 3.37. Let $r > 0$. Then by property (3) above and Example 3.33,

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \int_{\mathbb{R}^n} \frac{1}{|x|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,r)}(x)\,dx \\
&= r^n \int_{\mathbb{R}^n} \frac{1}{|rx|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,r)}(rx)\,dx \\
&= r^{n-\alpha} \int_{\mathbb{R}^n} \frac{1}{|x|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,1)}(x)\,dx \\
&= r^{n-\alpha} \int_{\mathbb{R}^n\setminus B(0,1)} \frac{1}{|x|^\alpha}\,dx < \infty, \quad \alpha > n,
\end{aligned}$$

and, in a similar way,

$$\int_{B(0,r)} \frac{1}{|x|^\alpha}\,dx = r^{n-\alpha}\int_{B(0,1)} \frac{1}{|x|^\alpha}\,dx < \infty, \quad \alpha < n.$$

Observe, that here we formally make the change of variables $x = ry$.

On the other hand, the integrals can be computed directly by (3.1). This gives

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \omega_{n-1}\int_r^\infty \rho^{-\alpha}\rho^{n-1}\,d\rho \\
&= \frac{\omega_{n-1}}{-\alpha+n}\rho^{-\alpha+n}\Big|_r^\infty = \frac{\omega_{n-1}}{\alpha-n}r^{-\alpha+n} < \infty, \quad \alpha > n
\end{aligned}$$

and

$$\begin{aligned}
\int_{B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \omega_{n-1}\int_0^r \rho^{-\alpha}\rho^{n-1}\,d\rho \\
&= \frac{\omega_{n-1}}{-\alpha+n}\rho^{-\alpha+n}\Big|_0^r = \frac{\omega_{n-1}}{n-\alpha}r^{n-\alpha} < \infty, \quad \alpha < n.
\end{aligned}$$

Remarks 3.38:

Formula (3.1) (or Example 3.33) implies following claims:

(1) If $|f(x)| \leqslant c|x|^{-\alpha}$ in a ball $B(0,r)$, $r > 0$, for some $\alpha < n$, then $f \in L^1(B(0,r))$. On the other hand, if $|f(x)| \geqslant c|x|^{-\alpha}$ in $B(0,r)$ for some $\alpha > n$, then $f \notin L^1(B(0,r))$.

(2) If $|f(x)| \leqslant c|x|^{-\alpha}$ in $\mathbb{R}^n\setminus B(0,r)$ for some $\alpha > n$, then $f \in L^1(\mathbb{R}^n\setminus B(0,r))$. On the other hand, if $|f(x)| \geqslant c|x|^{-\alpha}$ in $\mathbb{R}^n\setminus B(0,r)$ for some $\alpha < n$, then $f \notin L^1(\mathbb{R}^n\setminus B(0,r))$.

Approximation by continuous functions

An integrable function has the following approximation properties.

Theorem 3.39. Let $f \in L^1(\mathbb{R}^n)$ and $\varepsilon > 0$.

(1) There is a simple function $g \in L^1(\mathbb{R}^n)$ such that $\|f - g\|_{L^1(\mathbb{R}^n)} < \varepsilon$.

(2) There is a compactly supported continuous function $g \in C_0(\mathbb{R}^n)$ such that $\|f - g\|_{L^1(\mathbb{R}^n)} < \varepsilon$.

THE MORAL: Simple functions and compactly supported continuous functions are dense in $L^1(\mathbb{R}^n)$.

Proof. $\boxed{(1)}$ Since $f = f^+ - f^-$, we may consider f^+ and f^- separately and assume that $f \geqslant 0$. There is an increasing sequence of simple functions f_i, $i = 1, 2, \dots$, such that $f_i \to f$ everywhere in $\mathbb{R}^n$ as $i \to \infty$. By the dominated convergence theorem, we have

$$\lim_{i\to\infty} \|f_i - f\|_{L^1(\mathbb{R}^n)} = \lim_{i\to\infty} \int_{\mathbb{R}^n} |f_i - f|\,dx = \int_{\mathbb{R}^n} \underbrace{\lim_{i\to\infty} |f_i - f|}_{=0}\,dx = 0,$$

because $|f_i - f| \leqslant |f_i| + |f| \leqslant 2|f| \in L^1(\mathbb{R}^n)$ gives an integrable majorant.

$\boxed{(2)}$ **Step 1:** Since $f = f^+ - f^-$ we may assume that $f \geqslant 0$.

Step 2: The dominated convergence theorem gives

$$\begin{aligned}\lim_{i\to\infty} \|f\chi_{B(0,i)} - f\|_{L^1(\mathbb{R}^n)} &= \lim_{i\to\infty} \int_{\mathbb{R}^n} |f\chi_{B(0,i)} - f|\,dx \\ &= \int_{\mathbb{R}^n} \underbrace{\lim_{i\to\infty} |f\chi_{B(0,i)} - f|}_{=0}\,dx = 0,\end{aligned}$$

since $|f\chi_{B(0,i)} - f| \leqslant |f| \in L^1(\mathbb{R}^n)$. Thus compactly supported integrable functions are dense in $L^1(\mathbb{R}^n)$.

Step 3: By Theorem 2.32 there is an increasing sequence of simple functions $f_i : \mathbb{R}^n \to [0, \infty)$, $i = 1, 2, \dots$, such that $f_i \to f$ everywhere in $\mathbb{R}^n$ as $i \to \infty$. The dominated convergence theorem gives

$$\lim_{i\to\infty} \|f_i - f\|_{L^1(\mathbb{R}^n)} = \lim_{i\to\infty} \int_{\mathbb{R}^n} |f_i - f|\,dx = \int_{\mathbb{R}^n} \underbrace{\lim_{i\to\infty} |f_i - f|}_{=0}\,dx = 0.$$

since $|f_i - f| \leqslant |f_i| + |f| \leqslant 2|f| \in L^1(\mathbb{R}^n)$. Thus we can assume that we can approximate a nonnegative simple function which vanishes outside a bounded set.

Step 4: Such a function is of the form $f = \sum_{i=1}^k a_i \chi_{A_i}$, where A_i are bounded Lebesgue measurable set and $a_i \geqslant 0$, $i = 1, 2, \dots$. Thus if we can approximate each χ_{A_i} by a compactly supported continuous function, then the corresponding linear combination will approximate the simple function.

Step 5: Let A be a bounded Lebesgue measurable set and $\varepsilon > 0$. Since $m(A) < \infty$ by Theorem 1.59 there exist a compact set K and a open set G such that $K \subset A \subset G$ and $m(G \setminus K) < \varepsilon$.

Claim: There exists a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ such that

(1) $0 \leqslant g(x) \leqslant 1$ for every $x \in \mathbb{R}^n$,

(2) $g(x) = 1$ for every $x \in K$ and

(3) the support of g is a compact subset of G.

Reason. Let
$$U = \left\{x \in \mathbb{R}^n : \operatorname{dist}(x,K) < \tfrac{1}{2}\operatorname{dist}(K,\mathbb{R}^n \setminus G)\right\}.$$
Then $K \subset U \subset \overline{U} \subset G$, U is open and $\overline{U}$ is compact. The function $g : \mathbb{R}^n \to \mathbb{R}$,
$$g(x) = \frac{\operatorname{dist}(x,\mathbb{R}^n \setminus U)}{\operatorname{dist}(x,K) + \operatorname{dist}(x,\mathbb{R}^n \setminus U)},$$
has the desired properties, see Remark 1.29. Moreover,
$$\operatorname{supp} g = \overline{\{x \in \mathbb{R}^n : g(x) \neq 0\}} = \overline{U}$$
is compact. ■

Observe that
$$\begin{aligned}
&x \in K \Longrightarrow \chi_A(x) - g(x) = 1 - 1 = 0,\\
&x \in \mathbb{R}^n \setminus G \Longrightarrow \chi_A(x) - g(x) = 0 - 0 = 0,\\
&x \in A \setminus K \Longrightarrow \chi_A(x) - g(x) = 1 - g(x) < 1,\\
&x \in G \setminus A \Longrightarrow \chi_A(x) - g(x) = g(x) < 1.
\end{aligned}$$
Thus $|\chi_A - g| \leqslant 1$, $\chi_A - g$ vanishes in K and outside $G \setminus K$ and we have
$$\|\chi_A - g\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\chi_A - g|\,dx \leqslant m(G \setminus K) < \varepsilon.$$
This completes the proof of the approximation property. □

Remark 3.40. The claim (2) can also be proved using Lusin's theorem (Theorem 2.52) (exercise).

3.8 Cavalieri's principle

Recall, that every nonnegative measurable function satisfies Tchebyshev's inequality (Theorem 3.8 (5))
$$m(\{x \in \mathbb{R}^n : f(x) > t\}) \leqslant \frac{1}{t}\int_{\mathbb{R}^n} f\,dx, \quad t > 0.$$
In particular, if $f \in L^1(\mathbb{R}^n)$, then
$$m(\{x \in \mathbb{R}^n : f(x) > t\}) \leqslant \frac{c}{t}, \quad t > 0. \tag{3.2}$$
The converse claim is not true. That is, if f satisfies an inequality of the form (3.2), it does not follow that $f \in L^1(\mathbb{R}^n)$.

Reason. Let $f:\mathbb{R}^n \to [0,\infty]$, $f(x)=|x|^{-n}$. Then f satisfies (3.2), but $f \notin L^1(\mathbb{R}^n)$. ■

The function $t \mapsto m(\{x\in\mathbb{R}^n : f(x)>t\})$ is called the distribution function of f. Observe that the distribution set $\{x\in\mathbb{R}^n : f(x)>t\}$ is Lebesgue measurable and the distribution function is a nonincreasing function of $t>0$ and hence Lebesgue measurable. Let us consider the behaviour of $tm(\{x\in\mathbb{R}^n : f(x)>t\})$ as t increases. By Tchebyshev's inequality $tm(\{x\in\mathbb{R}^n : f(x)>t\}) \leqslant c$ for every $t>0$ if $f\in L^1(\mathbb{R}^n)$, but there is a stronger result.

Lemma 3.41. if $f\in L^1(\mathbb{R}^n)$ is a nonnegative function, then

$$\lim_{t\to\infty} tm(\{x\in\mathbb{R}^n : f(x)>t\})=0.$$

Proof. Let $A=\{x\in\mathbb{R}^n : f(x)<\infty\}$. Since $f\in L^1(\mathbb{R}^n)$, by Lemma 3.9 (2), we have $m(\mathbb{R}^n\setminus A)=0$. Denote $A_t=\{x\in\mathbb{R}^n : f(x)>t\}$, $t>0$. Then

$$A=\bigcup_{0<t<\infty}\mathbb{R}^n\setminus A_t \quad\text{and}\quad \lim_{t\to\infty}\chi_{\mathbb{R}^n\setminus A_t}(x)=\chi_A(x) \quad\text{for every}\quad x\in\mathbb{R}^n.$$

Clearly

$$\int_{\mathbb{R}^n} f\,dx=\int_{A_t} f\,dx+\int_{\mathbb{R}^n\setminus A_t} f\,dx.$$

By the dominated convergence theorem

$$\begin{aligned}\lim_{t\to\infty}\int_{\mathbb{R}^n\setminus A_t} f\,dx &= \lim_{t\to\infty}\int_{\mathbb{R}^n}\chi_{\mathbb{R}^n\setminus A_t} f\,dx\\ &=\int_{\mathbb{R}^n}\lim_{t\to\infty}\chi_{\mathbb{R}^n\setminus A_t} f\,dx=\int_{\mathbb{R}^n}\chi_A f\,dx\\ &=\int_A f\,dx=\int_A f\,dx+\underbrace{\int_{\mathbb{R}^n\setminus A} f\,dx}_{=0}=\int_{\mathbb{R}^n} f\,dx.\end{aligned}$$

Thus

$$\begin{aligned}\underbrace{\int_{\mathbb{R}^n} f\,dx}_{<\infty} &= \lim_{t\to\infty}\left(\int_{A_t} f\,dx+\int_{\mathbb{R}^n\setminus A_t} f\,dx\right)\\ &=\lim_{t\to\infty}\int_{A_t} f\,dx+\underbrace{\int_{\mathbb{R}^n} f\,dx}_{<\infty},\end{aligned}$$

which implies that

$$\lim_{t\to\infty}\int_{A_t} f\,dx=0.$$

By Tchebyshev's inequality

$$0\leqslant tm(A_t)\leqslant\int_{A_t} f\,dx\to 0 \quad\text{as}\quad t\to\infty,$$

which implies $tm(A_t)\to 0$ as $t\to\infty$. □

The following representation of the integral is very useful.

Theorem 3.42 (Cavalieri's principle). Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set and let $f : A \to [0,\infty]$ be a Lebesgue measurable function. Then

$$\int_A f\,dx = \int_0^\infty m(\{x \in A : f(x) > t\})\,dt.$$

T H E M O R A L : In order to estimate the integral of a function it is enough to estimate the distribution sets of the function.

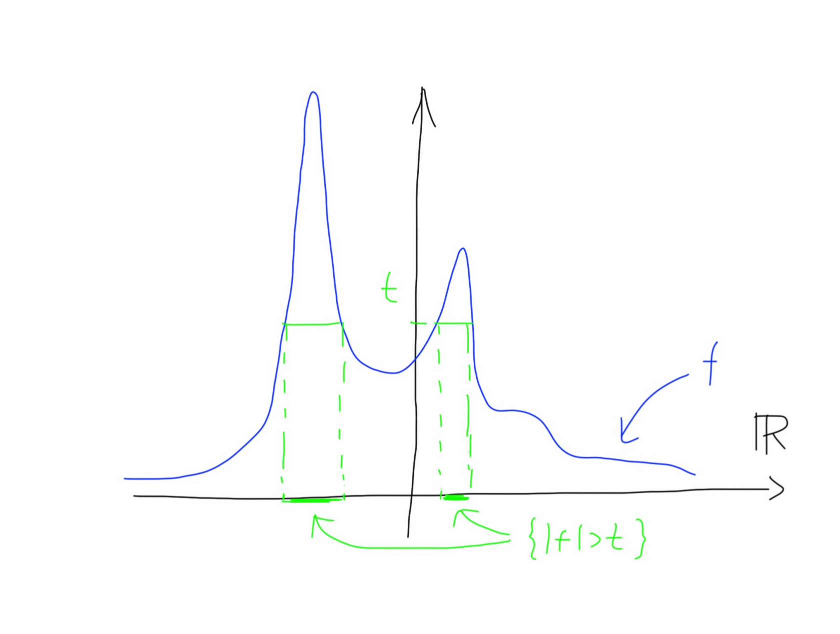

Figure 3.5: Cavelieri's principle.

Remarks 3.43:

(1) For signed Lebesgue measurable functions we have

$$\int_A |f|\,dx = \int_0^\infty m(\{x \in A : |f(x)| > t\})\,dt.$$

(2) It can be shown (without Cavalieri's principle) that

$$\int_A |f|^p\,dx < \infty \iff \sum_{i=-\infty}^{\infty} 2^{ip} m(\{x \in A : |f(x)| > 2^i\}) < \infty, \quad 0 < p < \infty.$$

(Exercise)

(3) If $g : A \to [0,\infty]$ is a rearrangement of f such that

$$m(\{x \in A : g(x) > t\}) = m(\{x \in A : f(x) > t\})$$

for every $t \geqslant 0$, then $\int_A g\,dx = \int_A f\,dx$.

(4) Cavalieri's principle can be taken as the definition of the Lebesgue integral but then we have to be able to define the right hand side of Cavalieri's principle without using the one-dimensional Lebesgue integral. If $\mu(A) < \infty$, then the distribution function is a bounded monotone function and thus continuous almost everywhere in $[0,\infty)$. This implies that the distribution function is Riemann integrable on any compact interval in $[0,\infty)$ and thus that the right-hand side of Cavalieri's principle can be interpreted as an improper Riemann integral, see Remark 3.47.

Proof. **Step 1:** First assume that f is a nonnegative simple function. Then $f = \sum_{i=0}^{k} a_i \chi_{A_i}$, where $A_i = f^{-1}(\{a_i\})$. We may assume that $0 = a_0 < a_1 < \cdots < a_k$. Then

$$\begin{aligned}
\int_0^\infty m(\{x \in A : f(x) > t\})\,dt &= \int_0^{a_k} m(\{x \in A : f(x) > t\})\,dt \\
&= \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} m(\{x \in A : f(x) > t\})\,dt \\
&= \sum_{i=1}^{k} (a_i - a_{i-1}) m\left(\bigcup_{j=i}^{k} A_j \cap A\right) \quad \left(f = \sum_{i=0}^{k} a_i \chi_{A_i}\right) \\
&= \sum_{i=1}^{k} (a_i - a_{i-1}) \sum_{j=i}^{k} m(A_j \cap A) \quad (A_j \text{ measurable and disjoint}) \\
&= \sum_{i=1}^{k} a_i \sum_{j=i}^{k} m(A_j \cap A) - \sum_{i=1}^{k} a_{i-1} \sum_{j=i}^{k} m(A_j \cap A) \\
&= \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=1}^{j} a_i - \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=0}^{j-1} a_i \\
&= \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=1}^{j} (a_i - a_{i-1}) = \sum_{j=1}^{k} a_j m(A_j \cap A) = \int_A f\,dx.
\end{aligned}$$

This proves the claim for nonnegative simple functions.

Step 2: Assume then that f is a nonnegative measurable function. Then there exists a sequence of nonnegative simple functions f_i, $i = 1,2,\ldots$, such that $f_i \leqslant f_{i+1}$ and $f(x) = \lim_{i\to\infty} f_i(x)$ for every $x \in A$. Thus

$$\{x \in A : f_i(x) > t\} \subset \{x \in A : f_{i+1}(x) > t\}$$

and

$$\bigcup_{i=1}^{\infty} \{x \in A : f_i(x) > t\} = \{x \in A : f(x) > t\}.$$

Denote

$$\varphi_i(t) = m(\{x \in A : f_i(x) > t\}) \quad \text{and} \quad \varphi(t) = m(\{x \in A : f(x) > t\}).$$

Then φ_j is an increasing sequence of functions and $\varphi_i(t) \to \varphi(t)$ for every $t \geqslant 0$ as

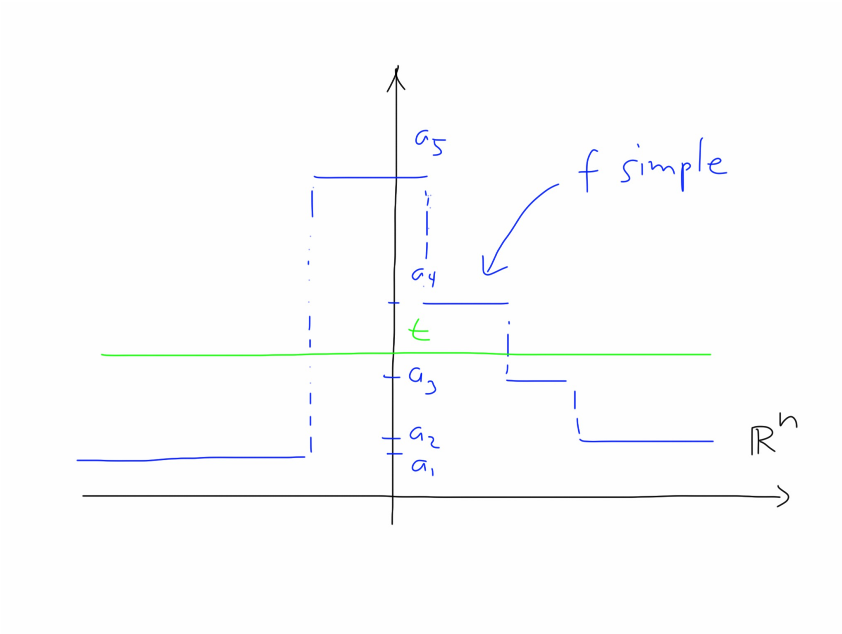

Figure 3.6: Cavalieri's principle for a simple function.

$i \to \infty$. The monotone convergence theorem implies

$$\begin{aligned}\int_0^\infty m(\{x \in A : f(x) > t\})\,dt &= \lim_{i\to\infty} \int_0^\infty m(\{x \in A : f_i(x) > t\})\,dt \\ &= \lim_{i\to\infty} \int_A f_i\,dx = \int_A f\,dx. \qquad \square\end{aligned}$$

Remarks 3.44:

(1) By a change of variables, we have

$$\int_A |f|^p\,dx = p \int_0^\infty t^{p-1} m(\{x \in A : |f(x)| > t\})\,dt$$

for $0 < p < \infty$.

(2) More generally, if $\varphi : [0,\infty) \to [0,\infty)$ is a nondecreasing continuously differentiable function with $\varphi(0) = 0$, then

$$\int_A \varphi \circ |f|\,dx = \int_0^\infty \varphi'(t) m(\{x \in A : |f(x)| > t\})\,dt.$$

(3) These results hold not only for the Lebesgue measure, but also for other measures.

We shall give another proof for Cavalieri's principle later, see Corollary 3.60.

Example 3.45. Let $f:\mathbb{R}^n \to \mathbb{R}$, $f(x)=|x|^{-\alpha}$, $0<\alpha<n$. Then

$$\begin{aligned}\int_{B(0,1)} f(x)\,dx &= \int_{B(0,1)} |x|^{-\alpha}\,dx = \int_0^\infty m(\{x\in B(0,1):|x|^{-\alpha}>t\})\,dt\\ &= \int_0^1 m(B(0,1))\,dt + \int_1^\infty m(\{x\in\mathbb{R}^n:|x|<t^{-1/\alpha}\})\,dt\\ &= m(B(0,1)) + \int_1^\infty m(B(0,t^{-1/\alpha}))\,dt\\ &= m(B(0,1)) + \int_1^\infty t^{-n/\alpha} m(B(0,1))\,dt\\ &= m(B(0,1))\Big(1+\frac{\alpha}{n-\alpha}\Big).\end{aligned}$$

3.9 Lebesgue and Riemann

THE MORAL: The main difference between the Lebesgue and Riemann integrals is that in the definition of the Riemann integral with step functions we subdivide the domain of the function but in the definition of the Lebesgue integral with simple functions we subdivide the range of the function.

We shall briefly recall the definition of the one-dimensional Riemann integral. Let I_i, $i=1,\dots,k$, be pairwise disjoint intervals in $\mathbb{R}$ with $\bigcup_{i=1}^k I_i=[a,b]$ with $a,b\in\mathbb{R}$ and let a_i, $i=1,\dots,k$, be real numbers. A function $f:[a,b]\to\mathbb{R}$ is said to be a step function, if

$$f=\sum_{i=1}^k a_i\chi_{I_i}.$$

Observe, that a step function is just a special type of a simple function. Let $f:[a,b]\to\mathbb{R}$ be a bounded function. Recall that the lower Riemann integral is

$$\underline{\int_a^b} f(x)\,dx = \sup\Big\{\int_{[a,b]} g\,dx : g\leqslant f \text{ on } [a,b] \text{ and } g \text{ is a step function}\Big\}$$

and the upper Riemann integral is

$$\overline{\int_a^b} f(x)\,dx = \inf\Big\{\int_{[a,b]} h\,dx : f\leqslant h \text{ on } [a,b] \text{ and } h \text{ is a step function}\Big\}.$$

Observe that we use the definition of the integral for a simple function for the integral of the step function. The function f is said to be Riemann integrable, if its lower and upper integrals coincide. The common value of lower and upper integrals is the Riemann integral of f on $[a,b]$ and it is denoted by

$$\int_a^b f(x)\,dx.$$

If $f:[a,b]\to\mathbb{R}$ is a bounded Lebesgue measurable function, then by the definition of the Lebesgue integral,

$$\underline{\int_a^b} f(x)\,dx \leqslant \int_{[a,b]} f(x)\,dx \leqslant \overline{\int_a^b} f(x)\,dx.$$

This implies that if f is Riemann integrable, then the Riemann and Lebesgue integrals coincide provided f is Lebegue measurable.

Lemma 3.46. Let $f:[a,b]\to\mathbb{R}$ be a bounded Riemann integrable function. Then f is Lebesgue integrable and

$$\int_a^b f(x)\,dx=\int_{[a,b]} f(x)\,dx.$$

THE MORAL: The Lebesgue integral is an extension of the Riemann integral.

Proof. By adding a constant we may assume that $f\geqslant 0$. By the definition of the Riemann integral, there are step functions g_i and h_i such that $g_i\leqslant f\leqslant h_i$,

$$\lim_{i\to\infty}\int_{[a,b]} g_i\,dx=\lim_{i\to\infty}\int_{[a,b]} h_i\,dx=\int_a^b f(x)\,dx.$$

By passing to sequences $\max\{g_1,\dots,g_j\}$ and $\min\{h_1,\dots,h_j\}$ we may assume that $g_i\leqslant g_{i+1}$ and $h_i\geqslant h_{i+1}$ for every $i=1,2,\dots$. These sequences are monotone and bounded and thus they converge pointwise. Denote

$$g(x)=\lim_{i\to\infty} g_i(x)\quad\text{and}\quad h(x)=\lim_{i\to\infty} h_i(x).$$

By the dominated convergence theorem

$$\int_a^b f(x)\,dx=\lim_{i\to\infty}\int_{[a,b]} g_i(x)\,dx=\int_{[a,b]} g(x)\,dx$$

and

$$\int_a^b f(x)\,dx=\lim_{i\to\infty}\int_{[a,b]} h_i(x)\,dx=\int_{[a,b]} h(x)\,dx.$$

Since $h-g\geqslant 0$ and

$$\begin{aligned}\int_{[a,b]}(h(x)-g(x))\,dx&=\int_{[a,b]} h(x)\,dx-\int_{[a,b]} g(x)\,dx\\&=\int_a^b f(x)\,dx-\int_a^b f(x)\,dx=0,\end{aligned}$$

we have $h-g=0$ almost everywhere in $[a,b]$. Since $g\leqslant f\leqslant h$ we have $h=g=f$ almost everywhere in $[a,b]$. Thus f is measurable and since it is also bounded it is integrable in $[a,b]$. $\square$

Remark 3.47. A necessary and sufficient condition for a function f to be Riemann integrable on an interval $[a,b]$ is that f is bounded and that its set of points of discontinuity in $[a,b]$ forms a set of Lebesgue measure zero. (Riemann-Lebesgue)

Note that the definition of the Riemann integral only applies to bounded functions defined on bounded intervals. It is possible to relax these assumptions, but this becomes delicate. The definition of the Lebesgue integral applies directly to not necessarily bounded functions and sets. Note that the Lebesgue integral is defined not only over intervals but also over more general measurable sets. This is a very useful property. Moreover, if $f_i \in L^1([a,b])$, $i = 1,2,\dots$, are functions such that

$$\sum_{i=1}^{\infty} \|f_i\|_{L^1([a,b])} < \infty,$$

then Corollary 3.16 implies that $f = \sum_{i=1}^{\infty} f_i$ is Lebesgue measurable. In addition, $f \in L^1([a,b])$ and

$$\int_{[a,b]} \sum_{i=1}^{\infty} f_i \, dx = \sum_{i=1}^{\infty} \int_{[a,b]} f_i \, dx.$$

The Riemann integral does not do very well here, since the limit function f can be discontinuous, for example, on a dense subset even if the functions f_i are continuous.

The following examples show differences between the Lebesgue and Riemann integrals.

Examples 3.48:

(1) $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \chi_{[0,1]\cap\mathbb{Q}}(x)$ is not Riemann integrable, but is a simple function for Lebesgue integral and

$$\int_{\mathbb{R}} \chi_{[0,1]\cap\mathbb{Q}}(x) \, dx = 0.$$

(2) Let q_i, $i = 1,2,\dots$, be an enumeration of rational numbers in the interval $[0,1]$ and define $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \chi_{\{q_1,\dots,q_i\}}(x), \quad i = 1,2,\dots$$

Then each f_i is Riemann integrable with zero integral, but the limit function

$$f(x) = \lim_{i\to\infty} f_i(x) = \chi_{[0,1]\cap\mathbb{Q}}(x)$$

is not Riemann integrable. This means that a pointwise limit of Riemann integrable functions may be not Riemann integrable.

(3) Define $f : [0,1] \to \mathbb{R}$ by setting $f(0) = 0$ and

$$f(x) = \begin{cases} \frac{2^{i+1}}{i}, & \frac{1}{2^i} < x \leqslant \frac{3}{2^{i+1}}, \\ -\frac{2^{i+1}}{i}, & \frac{3}{2^{i+1}} < x \leqslant \frac{1}{2^{i-1}}, \end{cases}$$

for $x \in (0,1]$. Note that $\frac{3}{2^{i+1}}$ is the midpoint of the interval $\left[\frac{1}{2^i}, \frac{1}{2^{i-1}}\right]$ and that the length of the interval is 2^{-i-1}. Then

$$\int_{[0,1]} f^+(x) \, dx = \sum_{i=1}^{\infty} \frac{2^{i+1}}{i} 2^{-i-1} = \sum_{i=1}^{\infty} \frac{1}{i} = \infty$$

and similarly

$$\int_{[0,1]} f^-(x)\,dx = \infty.$$

Thus f is not Lebesgue integrable in $[0,1]$. However, the improper integral

$$\int_0^1 f(x)\,dx = \lim_{\varepsilon\to 0}\int_\varepsilon^1 f(x)\,dx = 0$$

exists because of the cancellation.

(4) Let $f:[0,\infty)\to\mathbb{R}$

$$f(x)=\begin{cases}\frac{\sin x}{x}, & x>0,\\ 1, & x=0.\end{cases}$$

Observe that, since f is continuous, it is Lebesgue measurable.
Claim: $f\notin L^1([0,\infty))$.

Reason.

$$\begin{aligned}\int_{[0,\infty)}|f|\,dx &= \int_0^\pi |f(x)|\,dx + \sum_{i=1}^\infty \int_{i\pi}^{(i+1)\pi}\frac{|\sin x|}{|x|}\,dx\\ &\geqslant \int_0^\pi |f(x)|\,dx + \sum_{i=1}^\infty \frac{1}{i+1} = \infty.\end{aligned}$$

■

Claim: The improper Riemann integral $\lim_{a\to\infty}\int_0^a \frac{\sin x}{x}\,dx$ exits.

Reason. Denote $I(a)=\int_0^a \frac{\sin x}{x}\,dx$, $a\geqslant\pi$. Then

$$I(k\pi)=\int_0^{k\pi}\frac{\sin x}{x}\,dx = \sum_{i=0}^{k-1}\int_{i\pi}^{(i+1)\pi}\frac{\sin x}{x}\,dx,\quad k=1,2,\dots,$$

where

$$\sum_{i=0}^{k-1} a_i \quad\text{with}\quad a_i=\int_{i\pi}^{(i+1)\pi}\frac{\sin x}{x}\,dx,\quad i=1,2,\dots,$$

is an alternating series with the properties

$$a_i a_{i+1}<0,\quad |a_i|\leqslant|a_{i+1}|\quad\text{and}\quad \lim_{i\to\infty}a_i=0.$$

Thus this series converges and

$$s=\sum_{i=0}^\infty\int_{i\pi}^{(i+1)\pi}\frac{\sin x}{x}\,dx = \lim_{k\to\infty}I(k\pi).$$

Since $a\geqslant\pi$, we have $a\in[k\pi,(k+1)\pi)$ for some $k=1,2,\dots$ and

$$|I(a)-I(k\pi)|=\left|\int_{k\pi}^a\frac{\sin x}{x}\,dx\right|\leqslant\int_{k\pi}^a\frac{1}{k\pi}\,dx\leqslant\frac{1}{k}.$$

This shows that $\lim_{a\to\infty}I(a)=s$. ■

Thus f is not Lebesgue integrable on $[0,\infty)$, but the improper integral $\int_0^\infty f\,dx$ exists (and equals to $\frac{\pi}{2}$ by complex analysis).

Remark 3.49. There exists an everywhere differentiable function such that its derivative is bounded but not Riemann integrable. Let $C \subset [0,1]$ be a fat Cantor set with $m(C) = \frac{1}{2}$. Then

$$(0,1) \setminus C = \bigcup_{i=1}^{\infty} I_i,$$

where I_i are pairwise disjoint open intervals and $\sum_{i=1}^{\infty} \operatorname{vol}(I_i) = \frac{1}{2}$. For every $i = 1,2,\dots$, choose a closed centered subinterval $J_i \subset I_i$ such that $\operatorname{vol}(J_i) = \operatorname{vol}(I_i)^2$. Define a continuous function $f : [0,1] \to \mathbb{R}$ such that

$$f(x) = 0 \quad \text{for every} \quad x \in [0,1] \setminus \bigcup_{i=1}^{\infty} J_i,$$

$0 \leqslant f(x) \leqslant 1$ for every $x \in [0,1]$ and $f(x) = 1$ at the center of every J_i. The set $\bigcup_{i=1}^{\infty} I_i$ is dense in $[0,1]$, from which it follows that the upper Riemann integral is one and the lower Riemann integral is zero. Define

$$F(x) = \sum_{i=1}^{\infty} \int_{J_i \cap [0,x]} f(t)\,dt.$$

Then $F'(x) = f(x) = 0$ for every $x \in C$ (exercise) and $F'(x) = f(x)$ for every $x \in [0,1] \setminus C$. Thus f is a derivative.

3.10 Fubini's theorem

We shall show that certain multiple integrals can be computed as iterated integrals. Moreover, under certain assumptions, the value of an iterated integral is independent of the order of integration.

Definition 3.50. Let μ be an outer measure on X and ν an outer measure on Y. We define the product outer measure $\mu \times \nu$ on $X \times Y$ as

$$(\mu \times \nu)(S) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i) : S \subset \bigcup_{i=1}^{\infty} (A_i \times B_i) \right\},$$

where the infimum is taken over all collections of μ-measurable sets $A_i \subset X$ and ν-measurable sets $B_i \subset Y$, $i = 1,2,\dots$. The measure $\mu \times \nu$ is called the product measure of μ and ν.

Remark 3.51. It is an exercise to show that $\mu \times \nu$ is an outer measure on $X \times Y$.

THE MORAL: The product measure is defined in such a way that the "rectangles" inherit the correct measure $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$. Note that this holds true for all "rectangles" $A \times B$, where $A \subset X$ is a μ-measurable set and $B \subset Y$ is a ν-measurable set, see Fubini's theorem below.

Theorem 3.52 (Fubini's theorem). Let μ be an outer measure on X and ν an outer measure on Y.

(1) Then $\mu\times\nu$ is a regular outer measure on $X\times Y$, even if μ and ν are not regular.

(2) If $A\subset X$ is a μ-measurable set and $B\subset Y$ is a ν-measurable set, then $A\times B$ is $\mu\times\nu$-measurable and $(\mu\times\nu)(A\times B)=\mu(A)\nu(B)$.

(3) If $S\subset X\times Y$ is $\mu\times\nu$ measurable and both measures μ and ν are σ-finite, then $S_y=\{x\in X:(x,y)\in S\}$ is μ-measurable for ν-almost every $y\in Y$ and $S_x=\{y\in Y:(x,y)\in S\}$ is ν-measurable for ν-almost every $x\in X$. Moreover,

$$(\mu\times\nu)(S)=\int_Y \mu(S_y)\,d\nu(y)=\int_X \nu(S_x)\,d\mu(x).$$

(4) If f is $\mu\times\nu$-measurable, both measures μ and ν are σ-finite and the integral of f is defined (i.e. at least one of the functions f^+ and f^- is integrable) then

$$y\mapsto\int_X f(x,y)\,d\mu(x)$$

is a ν-measurable function,

$$x\mapsto\int_X f(x,y)\,d\nu(y)$$

is a μ-measurable function and

$$\begin{aligned}\int_{X\times Y} f(x,y)\,d(\mu\times\nu)&=\int_Y\left[\int_X f(x,y)\,d\mu(x)\right]d\nu(y)\\&=\int_X\left[\int_Y f(x,y)\,d\nu(y)\right]d\mu(x).\end{aligned}$$

THE MORAL: Claim (2) shows how to compute the product measure $(\mu\times\nu)(A\times B)$ of a rectangle by using the two measures $\mu(A)$ and $\nu(B)$. For a more general set S the product measure $(\mu\times\nu)(S)$ can be computed by integrating over the slices of the set S parallel to the coordinate axes by claim (3). Claim (4) shows that the finite integral of a function with respect to a product measure can be obtained from the two iterated integrals.

Proof. $\boxed{(1)}$ Let $\mathscr{F}$ denote the collection of all sets $S\subset X\times Y$ for which the integral

$$\int_X \chi_S(x,y)\,d\mu(x)$$

exists for ν-almost every $y\in Y$ and, in addition, such that

$$\rho(S)=\int_Y\left[\int_X \chi_S(x,y)\,d\mu(x)\right]d\nu(y)$$

exists. Note that $+\infty$ is allowed here.

Claim: If $S_i\in\mathscr{F}$, $i=1,2,\dots$, are pairwise disjoint, then $S=\bigcup_{i=1}^{\infty}S_i\in\mathscr{F}$.

Reason. Note that $\chi_S = \sum_{i=1}^{\infty} \chi_{S_i}$. By Corollary 3.16 we have $\rho(S) = \sum_{i=1}^{\infty} \rho(S_i)$. This shows that $\mathscr{F}$ is closed under countable unions of pairwise disjoint sets. ∎

Claim: If $S_i \in \mathscr{F}$, $i = 1,2,\dots$, $S_1 \supset S_2 \supset \dots$, and $\rho(S_1) < \infty$, then $S = \bigcap_{i=1}^{\infty} S_i \in \mathscr{F}$.

Reason. Note that $\chi_S = \lim_{i\to\infty} \chi_{S_i}$. By the dominated convergence theorem we have

$$\rho(S) = \lim_{i\to\infty} \rho(S_i).$$

This shows that $\mathscr{F}$ is closed under decreasing convergence of sets with a finiteness condition. ∎

Define

$$\mathscr{P}_0 = \{A \times B : A \text{ is } \mu\text{-measurable and } B \text{ is } \nu\text{-measurable}\},$$
$$\mathscr{P}_1 = \left\{ \bigcup_{i=1}^{\infty} S_i : S_i \in \mathscr{P}_0 \right\} \quad \text{and} \quad \mathscr{P}_2 = \left\{ \bigcap_{i=1}^{\infty} S_i : S_i \in \mathscr{P}_1 \right\}.$$

The members of $\mathscr{P}_0$ are called measurable rectangles, the class $\mathscr{P}_1$ consists of countable unions of measurable rectangles and and $\mathscr{P}_2$ of countable intersections of these. The latter sets constitute a class relative to which the product measure will be regular.

Note that $\mathscr{P}_0 \subset \mathscr{F}$ and

$$\rho(A \times B) = \mu(A)\nu(B)$$

whenever $A \times B \in \mathscr{P}_0$. If $A_1 \times B_1, A_2 \times B_2 \in \mathscr{P}_0$, then

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2) \in \mathscr{P}_0$$

and

$$(A_1 \times B_1) \setminus (A_2 \times B_2) = ((A_1 \setminus A_2) \times B_1) \cup ((A_1 \cap A_2) \times (B_1 \setminus B_2)) \in \mathscr{P}_0$$

as a disjoint union of members of $\mathscr{P}_0$. As in the proof of Theorem 1.9 it follows that every member of $\mathscr{P}_1$ is a countable union of pairwise disjoint members of $\mathscr{P}_0$ and hence $\mathscr{P}_1 \subset \mathscr{F}$. $\mathscr{P}_2$

$\boxed{(2)}$ **Claim:** For every $S \subset X \times Y$,

$$(\mu \times \nu)(S) = \inf\{\rho(R) : S \subset R \in \mathscr{P}_1\}.$$

Reason. Suppose that $A_i \times B_i \in \mathscr{P}_0$, $i = 1,2,\dots$ and $S \subset R = \bigcup_{i=1}^{\infty}(A_i \times B_i)$. Then $\chi_R \leqslant \sum_{i=1}^{\infty} \chi_{A_i \times B_i}$ and

$$\rho(R) \leqslant \sum_{i=1}^{\infty} \rho(A_i \times B_i) = \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i).$$

Thus

$$\inf\{\rho(R) : S \subset R \in \mathscr{P}_1\} \leqslant (\mu \times \nu)(S).$$

Moreover, if $R = \bigcup_{i=1}^{\infty}(A_i \times B_i)$ is any such set, there exist pairwise disjoint sets $A_i' \times B_i' \in \mathscr{P}_0$ such that

$$R = \bigcup_{i=1}^{\infty}(A_i \times B_i) = \bigcup_{i=1}^{\infty}(A_i' \times B_i').$$

Thus

$$\rho(R) = \sum_{i=1}^{\infty} \mu(A_i')\nu(B_i') \geqslant (\mu \times \nu)(S).$$

This shows that the equality holds. ■

$\boxed{(3)}$ Fix $A \times B \in \mathscr{P}_0$. Then

$$(\mu \times \nu)(A \times B) \leqslant \mu(A)\nu(B) = \rho(A \times B) \leqslant \rho(R)$$

for every $R \in \mathscr{P}_1$ such that $A \times B \subset R$. Thus the claim above implies

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

We show that $A \times B$ is $\mu \times \nu$-measurable. Let $R \in \mathscr{P}_1$ with $T \subset R$. Then $R \setminus (A \times B)$ and $R \cap (A \times B)$ are disjoint members of $\mathscr{P}_1$. Thus

$$\begin{aligned}(\mu \times \nu)&(T \setminus (A \times B)) + (\mu \times \nu)(T \cap (A \times B)) \\ &\leqslant \rho(R \setminus (A \times B)) + \rho(R \cap (A \times B)) = \rho(R).\end{aligned}$$

The claim above implies

$$(\mu \times \nu)(T \setminus (A \times B)) + (\mu \times \nu)(T \cap (A \times B)) \leqslant (\mu \times \nu)(T).$$

Since this holds for all $T \subset X \times Y$, we have shown that $A \times B$ is $\mu \times \nu$-measurable. Theorem 1.9 implies that $\mathscr{P}_0$, $\mathscr{P}_1$ and $\mathscr{P}_2$ consist of $\mu \times \nu$-measurable sets. This proves claim (2) of the theorem.

$\boxed{(4)}$ Next we show that $\mu \times \nu$ is a regular measure.

Claim: For every $S \subset X \times Y$ there is $R \in \mathscr{P}_2$ such that $S \subset R$ and

$$(\mu \times \nu)(S) = (\mu \times \nu)(R) = \rho(R).$$

Reason. If $(\mu \times \nu)(S) = \infty$, set $R = X \times Y$. Thus we may assume that $(\mu \times \nu)(S) < \infty$. By the claim in (2), for every $i = 1, 2, \dots$ there exists a set $R_i \in \mathscr{P}_1$ such that $S \subset R_i$ and

$$\rho(R_i) < (\mu \times \nu)(S) + \frac{1}{i}.$$

Let $R = \bigcap_{i=1}^{\infty} R_i \in \mathscr{P}_2$. Since $R_i \in \mathscr{F}$ for every $i = 1, 2, \dots$, we conclude that $R \in \mathscr{F}$ and by the dominated convergence theorem

$$(\mu \times \nu)(S) \leqslant \rho(R) = \lim_{k\to\infty} \rho\left(\bigcap_{i=1}^{k} R_i\right) \leqslant (\mu \times \nu)(S).$$

This show that $\mu \times \nu$ in $\mathscr{P}_2$-regular. The claim follows from this, since every set in $\mathscr{P}_2$ is $\mu \times \nu$-measurable by claim 2 of the theorem. ■

$\boxed{(5)}$ If $S \subset X \times Y$ with $(\mu \times \nu)(S) = 0$, then there exists a set $R \in \mathscr{P}_2$ such that $S \subset R$ and $\rho(R) = 0$. Thus $S \in \mathscr{F}$ and $\rho(S) = 0$.

Suppose that $S \subset X \times Y$ is $\mu \times \nu$-measurable and $(\mu \times \nu)(S) < \infty$. Then there is $R \in \mathscr{P}_2$ such that $S \subset R$ and $(\mu \times \nu)(R \setminus S) = 0$. and, consequently, $\rho(R \setminus S) = 0$. It follows that

$$\mu(\{x \in X : (x, y) \in S\}) = \mu(\{x \in X : (x, y) \in R\})$$

for ν-almost every $y \in Y$ and

$$(\mu \times \nu)(S) = \rho(R) = \int_Y \mu(\{x \in X : (x, y) \in S\})\, d\nu(y).$$

This proves claim (3) of the theorem, because the other formula is symmetric with X replaced by Y and μ by ν. The extension to σ-finite $\mu \times \nu$-measure is obvious.

$\boxed{(6)}$ Claim (4) reduces to (3) when $f = \chi_S$. If f is a nonnegative $\mu \times \nu$-measurable function and is σ-finite with respect to $\mu \times \nu$, we use approximation by simple functions and the monotone convergence theorem. Finally, for general f we write $f = f^+ - f^-$. □

Following Tonelli's theorem for nonnegative product measurable functions is a corollary of Fubini's theorem, but it is useful to restate it in this form.

Theorem 3.53 (Tonelli's theorem). Let μ be an outer measure on X and ν an outer measure on Y and suppose that both measures are σ-finite. Let $f : X \times Y \to [0, \infty]$ be a nonnegative $\mu \times \nu$-measurable function. Then

$$y \mapsto \int_X f(x, y)\, d\mu(x)$$

is a ν-measurable function,

$$x \mapsto \int_X f(x, y)\, d\nu(y)$$

is a μ-measurable function and

$$\begin{aligned}\int_{X\times Y} f(x, y)\, d(\mu \times \nu) &= \int_Y \left[\int_X f(x, y)\, d\mu(x)\right] d\nu(y)\\ &= \int_X \left[\int_Y f(x, y)\, d\nu(y)\right] d\mu(x).\end{aligned}$$

THE MORAL: The order of iterated integrals can be switched for all nonnegative product measurable functions even in the case when the integrals are infinite.

Remarks 3.54:

(1) If μ and ν are counting measures, Tonelli's theorem reduces to a corresponding claim for series. Let $x_{i,j} \in [0,\infty]$, $i,j = 1,2,\dots$. Then

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_{i,j} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} x_{i,j}.$$

This means that we may rearrange the series without affecting the sum if the terms are nonnegative, compare with Corollary 3.16.

(2) Let μ be an outer measure on X and ν an outer measure on Y and suppose that both measures are σ-finite. Let $f : X \to [-\infty,\infty]$ be a $\mu\times\nu$-measurable function. If any of the three integrals

$$\begin{gathered}\int_{X\times Y} |f(x,y)|\,d(\mu\times\nu),\\ \int_Y \left[\int_X |f(x,y)|\,d\mu(x)\right] d\nu(y),\\ \int_X \left[\int_Y |f(x,y)|\,d\nu(y)\right] d\mu(x)\end{gathered}$$

is finite, then all of them are finite and the conclusion of Fubini's theorem holds (exercise). In particular, it follows that the function $y \mapsto f(x,y)$ is ν-integrable for μ-almost every $x \in X$ and that the function $x \mapsto f(x,y)$ is μ-integrable for ν-almost every $y \in Y$.

3.11 Fubini's theorem for Lebesgue measure

We shall reformulate Tonelli's and Fubini's theorems for the Lebesgue measure, see [3, Section 7.4], [7, Chapter 8], [11, p. 75–86] and [14, Chapter 6]. If $A \subset \mathbb{R}^n$ is a m^n-measurable set and $B \subset \mathbb{R}^m$ is a m^m-measurable set, then $A\times B \subset \mathbb{R}^{n+m}$ is a m^{n+m}-measurable set and

$$m^{n+m}(A\times B) = (m^n \times m^m)(A\times B) = m^n(A)m^m(B),$$

with the understanding that if one of the sets is of measure zero, then the product set is of measure zero. The connection between the $(n+m)$-dimensional Lebesgue measure m^{n+m} and the product measure $m^n \times m^m$ is delicate. The key result is that the Lebesgue measure on $\mathbb{R}^n \times \mathbb{R}^m$ is the completion of the product of the Lebesgue measures on $\mathbb{R}^n$ and $\mathbb{R}^m$, see [3, Theorem 7.29].

Theorem 3.55 (Tonelli's theorem). Let $f : \mathbb{R}^{n+m} \to [0,\infty]$ be a nonnegative m^{n+m}-measurable function. Then $y \mapsto f(x,y)$ is m^m-measurable for m^n-almost every $x \in \mathbb{R}^n$, $x \mapsto f(x,y)$ is m^n-measurable for m^m-almost every $y \in \mathbb{R}^m$,

$$y \mapsto \int_X f(x,y)\,dm^n(x)$$

is a m^m-measurable function

$$x \mapsto \int_X f(x,y)\,dm^m(y)$$

is a m^n-measurable function and

$$\begin{aligned}\int_{\mathbb{R}^{n+m}} f(x,y)\,dm^{n+m} &= \int_{\mathbb{R}^m}\left[\int_{\mathbb{R}^n} f(x,y)\,dm^n(x)\right]dm^m(y)\\ &= \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^m} f(x,y)\,dm^m(y)\right]dm^n(x).\end{aligned}$$

Remarks 3.56:

(1) The function $x \mapsto f(x,y)$ is not necessary a m^n-measurable function for every $y \in \mathbb{R}^m$. Nor is the slice $A_y = \{x \in \mathbb{R}^n : (x,y) \in A\}$ a m^n-measurable set for every $y \in \mathbb{R}^m$. Let $E \subset \mathbb{R}$ be a set which is not m^1-measurable and consider $A = E \times \{0\} = \{(x,y) \in \mathbb{R}^2 : x \in E,\, y = 0\}$. Then $m^2(A) = 0$ and thus A is m^2-measurable, but A_y is not m^1-measurable for $y = 0$. Note that measurability holds for almost every slice.

(2) If A is m^{n+m}-measurable, then the slice $A_y = \{x \in \mathbb{R}^n : (x,y) \in A\}$ is m^n-measurable for m^m-almost every $y \in \mathbb{R}^m$. A corresponding statement holds with the roles of x and y interchanged. Let $E \subset \mathbb{R}$ be a set which is not m^1-measurable and consider $A = [0,1] \times E \subset \mathbb{R} \times \mathbb{R}$. Then

$$A_y = \begin{cases} [0,1], & y \in E,\\ \emptyset, & y \notin E.\end{cases}$$

Thus A_y is m^1-measurable for every $y \in \mathbb{R}$. However, if A were m^2-measurable, then $A_x = \{y \in \mathbb{R} : (x,y) \in A\}$ is m^1-measurable for almost every $x \in \mathbb{R}$. This is not true, since $A_x = E$ for every $x \in [0,1]$. There exists a set $A \subset [0,1] \times [0,1]$, which is not m^2-measurable with the property that A_y and A_x are m^1-measurable for every $x, y \in [0,1]$ with $m^1(A_y) = 0$ and $m^1(A_x) = 1$ for every $x, y \in [0,1]$, see [11, p. 82-83].

Theorem 3.57 (Fubini's theorem). Let $f : \mathbb{R}^{n+m} \to [-\infty,\infty]$ be a m^{n+m}-measurable function and suppose that at least one of the integrals

$$\begin{aligned}&\int_{\mathbb{R}^{n+m}} |f(x,y)|\,dm^{n+m},\\ &\int_{\mathbb{R}^m}\left[\int_{\mathbb{R}^n} |f(x,y)|\,dm^n(x)\right]dm^m(y),\\ &\int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^m} |f(x,y)|\,dm^m(y)\right]dm^n(x),\end{aligned}$$

is finite. Then $y \mapsto f(x,y)$ is integrable in $\mathbb{R}^m$ for m^n-almost every $x \in \mathbb{R}^n$, $x \mapsto f(x,y)$ is integrable in $\mathbb{R}^n$ for m^m-almost every $y \in \mathbb{R}^m$,

$$y \mapsto \int_{\mathbb{R}^n} f(x,y)\,dm^n(x)$$

is integrable in $\mathbb{R}^m$,

$$x \mapsto \int_{\mathbb{R}^m} f(x,y)\,dm^m(y)$$

is integrable in $\mathbb{R}^n$ and

$$\begin{aligned}\int_{\mathbb{R}^{n+m}} f(x,y)\,dm^{n+m} &= \int_{\mathbb{R}^m}\left[\int_{\mathbb{R}^n} f(x,y)\,dm^n(x)\right] dm^m(y)\\ &= \int_{\mathbb{R}^n}\left[\int_{\mathbb{R}^m} f(x,y)\,dm^m(y)\right] dm^n(x).\end{aligned}$$

We consider few corollaries of the previous theorems.

Corollary 3.58. Suppose that $f:\mathbb{R}^n \to [-\infty,\infty]$ is a m^n-measurable function. Then the function $\widetilde{f}:\mathbb{R}^m \to [-\infty,\infty]$ defined by $\widetilde{f}(x,y) = f(x)$ is a m^{n+m}-measurable function.

Proof. We may assume that f is real valued. Since f is m^n-measurable, the set $A = \{x \in \mathbb{R}^n : f(x) < a\}$ is a m^n-measurable for every $a \in \mathbb{R}$. Since

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : \widetilde{f}(x,y) < a\} = A \times \mathbb{R}^m,$$

we conclude that the set is m^{n+m}-measurable for every $a \in \mathbb{R}$. Thus $\widetilde{f}$ is a m^{n+m}-measurable function. □

Corollary 3.59. Assume that $f:\mathbb{R}^n \to [0,\infty]$ is a nonnegative function and let

$$A = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : 0 \leqslant y \leqslant f(x)\}.$$

Then the following claims are true:

(1) f is a m^n-measurable function if and only if A is a m^{n+1}-measurable set.

(2) If the conditions in (1) hold, then

$$\int_{\mathbb{R}^n} f\,dm^n = m^{n+1}(A).$$

THE MORAL: The integral describes the area under the graph of a function.

Proof. Assume that f is a m^n-measurable function. By Corollary 3.58 the functions $(x,y) \mapsto -f(x)$ and $(x,y) \mapsto y$ are m^{n+1}-measurable and thus

$$F(x,y) = y - f(x)$$

is m^{n+1}-measurable. This implies that

$$A = \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : y \geqslant 0\} \cap \{(x,y) \in \mathbb{R}^n \times \mathbb{R} : F(x,y) \leqslant 0\}$$

is m^{n+1}-measurable.

Conversely, suppose that A is m^{n+1}-measurable. For every $x \in \mathbb{R}^n$ the slice

$$A_x = \{y \in \mathbb{R} : (x,y) \in A\} = [0, f(x)]$$

is a closed one-dimensional interval. By Fubini's theorem $m^1(A_x) = f(x)$ is a measurable function and

$$\begin{aligned}
m^{n+1}(A) &= \int_{\mathbb{R}^{n+1}} \chi_A(x,y)\, dm^{n+1}(x,y) \\
&= \int_{\mathbb{R}^n} m(A_x)\, dm^n(x) \\
&= \int_{\mathbb{R}^n} f(x)\, dm^n(x). \qquad \square
\end{aligned}$$

We give an alternative proof for Cavalieri's principle by Fubini's theorem. Compare to Theorem 3.42.

Corollary 3.60. Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set and let $f : A \to [0,\infty]$ be a Lebesgue measurable function. Then

$$\int_A f\, dx = \int_0^\infty m(\{x \in A : f(x) > t\})\, dt.$$

Proof.

$$\begin{aligned}
\int_A |f|\, dx &= \int_{\mathbb{R}^n} \chi_A(x) f(x)\, dx \\
&= \int_{\mathbb{R}^n} \int_0^\infty \chi_A(x) \chi_{[0,f(x))}(t)\, dt\, dx \\
&= \int_0^\infty \int_{\mathbb{R}^n} \chi_A(x) \chi_{[0,f(x))}(t)\, dx\, dt \quad \text{(Fubini)} \\
&= \int_0^\infty \int_{\mathbb{R}^n} \chi_A(x) \chi_{\{x \in \mathbb{R}^n : f(x) > t\}}(x)\, dx\, dt \\
&= \int_0^\infty m(\{x \in A : f(x) > t\})\, dt. \qquad \square
\end{aligned}$$

Example 3.61. Suppose that $A \subset \mathbb{R}^2$ is a Lebesgue measurable set with $m^2(A) = 0$. We claim that almost every horizontal line intersects A in a set whose one-dimensional Lebesgue measure is zero. The corresponding claim holds for vertical lines as well.

Reason. Let $A_1(y) = \{x \in \mathbb{R} : (x,y) \in A\}$ and $A_2(x) = \{y \in \mathbb{R} : (x,y) \in A\}$ with $x, y \in \mathbb{R}$. We shall show that $m^1(A_1(y)) = 0$ for almost every $y \in \mathbb{R}$ and, correspondingly, $m^1(A_2(x)) = 0$ for almost every $x \in \mathbb{R}$. Let $f = \chi_A$. Fubini's theorem implies

$$0 = m^2(A) = \int_{\mathbb{R}^2} \chi_A\, dm^2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x,y)\, dy \right) dx.$$

It follows that

$$m^1(A_2(x)) = \int_{\mathbb{R}} f(x,y)\,dy = 0$$

for almost every $x \in \mathbb{R}$. The proof for the claim $m^1(A_1(y)) = 0$ for almost every $y \in \mathbb{R}$ is analogous. ■

Conversely, if $A \subset \mathbb{R}^2$ such that $m^1(A_1(y)) = 0$ for almost every $y \in \mathbb{R}$ or $m^1(A_2(x)) = 0$ for almost every $x \in \mathbb{R}$, then $m^2(A) = 0$.

Reason. Fubini's theorem for the measurable function $f = \chi_A$. ■

WARNING: The assumption that $A \subset \mathbb{R}^2$ is measurable is essential. Indeed, there exist a set $A \subset \mathbb{R}^2$ such that

(1) A is not Lebesgue measurable and thus $m(A) > 0$,

(2) every horizontal line intersects A at most one point and

(3) every vertical line intersects A at most one point.

(Sierpinski: Fundamenta Mathematica 1 (1920), p. 114)

The examples below show how we can use Fubini's theorem to evaluate multiple integrals.

Examples 3.62:

(1) We show that

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,dx = \sqrt{\pi}$$

in two ways.

$\boxed{(1)}$ Note that $\mathrm{e}^{-x^2} > 0$ for every $x \in \mathbb{R}$ and

$$I \leqslant \int_{-\infty}^{-1} -x\mathrm{e}^{-x^2}\,dx + \int_{-1}^{1} \mathrm{e}^{-x^2}\,dx + \int_{1}^{\infty} x\mathrm{e}^{-x^2}\,dx < \infty.$$

Since $x \mapsto \mathrm{e}^{-x^2}$ is even,

$$I = 2\int_0^{\infty} \mathrm{e}^{-x^2}\,dx.$$

Now

$$I^2 = \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2}\,dx\right)\left(\int_{-\infty}^{\infty} \mathrm{e}^{-y^2}\,dy\right) = 4\int_0^{\infty}\left(\int_0^{\infty} \mathrm{e}^{-(x^2+y^2)}\,dy\right)dx.$$

Substitution $y = xs$ implies $dy = x\,ds$. By Fubini's theorem

$$\begin{aligned}\frac{I^2}{4} &= \int_0^{\infty}\left(\int_0^{\infty} \mathrm{e}^{-(1+s^2)x^2}\,x\,ds\right)dx = \int_0^{\infty}\left(\int_0^{\infty} \mathrm{e}^{-(1+s^2)x^2}\,x\,dx\right)ds \\ &= \frac{1}{2}\int_0^{\infty} \frac{1}{1+s^2}\,ds = \frac{1}{2}\arctan s\Big|_0^{\infty} = \frac{\pi}{4}.\end{aligned}$$

Thus $I = \sqrt{\pi}$.

$\boxed{(2)}$

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy &= \lim_{i\to\infty}\int_{B(0,i)} e^{-(x^2+y^2)}\,dx\,dy = \lim_{i\to\infty}\int_{B(0,i)} e^{-(x^2+y^2)}\,dx\,dy\\ &= \lim_{i\to\infty}\int_0^i\int_0^{2\pi} e^{-r^2}\,r\,dr\,d\theta = 2\pi\lim_{i\to\infty}\int_0^i e^{-r^2}\,r\,dr\\ &= \pi\lim_{i\to\infty} -e^{-r^2}\Big|_0^i = -\pi\lim_{i\to\infty}(e^{-i^2}-1) = \pi,\end{aligned}$$

where all integrals are (possibly inproper) Riemann integrals. We used a change of variables to the polar coordinates. By the Lebesgue monotone convergence theorem the Riemann and Lebesgue integrals

$$\int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy$$

coincide. By Fubini's theorem

$$\begin{aligned}\pi &= \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\,dx\,dy = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} e^{-x^2}e^{-y^2}\,dy\right)dx\\ &= \int_{\mathbb{R}} e^{-x^2}\left(\int_{\mathbb{R}} e^{-y^2}\,dy\right)dx = \left(\int_{\mathbb{R}} e^{-x^2}\,dx\right)^2\end{aligned}$$

and thus

$$\int_{\mathbb{R}} e^{-x^2}\,dx = \sqrt{\pi}.$$

(2) Consider

$$\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\,dx, \quad a,b>0.$$

Since

$$\frac{e^{-ax}-e^{-bx}}{x} = \int_a^b e^{-xy}\,dy,$$

we have

$$\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\,dx = \int_0^\infty\int_a^b e^{-xy}\,dy\,dx.$$

The function $(x,y)\mapsto e^{-xy}$ is continuous and thus Lebesgue measurable. Since $e^{-xy}>0$, we have

$$\int_0^\infty \frac{e^{-ax}-e^{-bx}}{x}\,dx = \int_a^b\int_0^\infty e^{-xy}\,dx\,dy = \int_a^b \frac{1}{y}\,dy = \log\frac{b}{a}$$

The following examples shows that we have to be careful when we apply Fubini's theorem.

Examples 3.63:

(1) Consider

$$\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\,dx.$$

Note that

$$\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy = \int_0^1 \frac{x^2+y^2}{(x^2+y^2)^2}\,dy + \int_0^1 \frac{-2y^2}{(x^2+y^2)^2}\,dy$$
$$= \int_0^1 \frac{1}{x^2+y^2}\,dy + \int_0^1 y\left(\frac{d}{dy}\frac{1}{x^2+y^2}\right)dy.$$

An integration by parts gives

$$\int_0^1 y\left(\frac{d}{dy}\frac{1}{x^2+y^2}\right)dy = \left.\frac{y}{x^2+y^2}\right|_0^1 - \int_0^1 \frac{1}{x^2+y^2}\,dy$$
$$= \frac{1}{x^2+1} - \int_0^1 \frac{1}{x^2+y^2}\,dy.$$

Thus

$$\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy = \frac{1}{x^2+1},$$

from which it follows that

$$\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\,dx = \int_0^1 \frac{1}{x^2+1}\,dx = \arctan x\Big|_0^1 = \frac{\pi}{4}.$$

By symmetry

$$\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dx\,dy = -\int_0^1\int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\,dx = -\frac{\pi}{4}.$$

Observe that in this case both iterated integrals exist and are finite, but they are not equal. This does not contradict Fubini's theorem, since

$$\int_0^1\int_0^1 \left|\frac{x^2-y^2}{(x^2+y^2)^2}\right| dy\,dx = \int_0^1 \left[\int_0^x \frac{x^2-y^2}{(x^2+y^2)^2}\,dy + \int_x^1 \frac{y^2-x^2}{(x^2+y^2)^2}\,dy\right] dx$$
$$= \int_0^1 \frac{1}{x}\,dx - \int_0^1 \frac{1}{x^2+1}\,dx = \infty.$$

The integral in the brackets can be evaluated by integration by parts.

(2) Let

$$A = \bigcup_{i=1}^{\infty} \chi_{[i,i+1]\times[i,i+1]} \quad\text{and}\quad B = \bigcup_{i=1}^{\infty} \chi_{[i+1,i+2]\times[i,i+1]}.$$

Consider $f:\mathbb{R}^2\to\mathbb{R}$, $f=\chi_A-\chi_B$. Then

$$\int_{[0,1]} f(x,y)\,dx = 0$$

for every $y\in\mathbb{R}$ and

$$\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(x,y)\,dx\right)dy = 0.$$

On the other hand,

$$\int_{[0,1]} f(x,y)\,dy = \chi_{[0,1]}(x)$$

for every $x\in\mathbb{R}$. Thus

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

(3) Let $f:[0,1]\times[0,1]\to\mathbb{R}$ be defined by

$$f(x,y)=\begin{cases}2^{2i}, & 2^{-i}\leqslant x<2^{-i+1},\, 2^{-i}\leqslant y<2^{-i+1},\\ -2^{2i+1}, & 2^{-i-1}\leqslant x<2^{-i},\, 2^{-i}\leqslant y<2^{-i+1},\\ 0, & \text{otherwise.}\end{cases}$$

Then the integral of f over any horizontal line in the left half of the square is 0 and the integral over any vertical line in the right half of the square is 2. In this case

$$\begin{aligned}&\int_{[0,1]}\left(\int_{[0,1]} f(x,y)\,dx\right)dy=0,\\ &\int_{[0,1]}\left(\int_{[0,1]} f(x,y)\,dy\right)dx=1 \quad\text{and}\\ &\int_{[0,1]}\left(\int_{[0,1]} |f(x,y)|\,dy\right)dx=\infty.\end{aligned}$$

(Exercise)

(4) Let $0=\delta_1<\delta_2<\ldots\cdots<1$ and $\delta_i\to 1$ as $i\to\infty$. Let g_i, $i=1,2,\ldots$ be continuous functions such that

$$\operatorname{supp} g_i\subset(\delta_i,\delta_{i+1}) \quad\text{and}\quad \int_0^1 g_i(t)\,dt=1$$

for every $i=1,2,\ldots$. Define

$$f(x,y)=\sum_{i=1}^{\infty}[g_i(x)-g_{i+1}(x)]g_i(y).$$

The function f is continuous except at the point $(1,1)$, but

$$\int_0^1\int_0^1 f(x,y)\,dy\,dx=1\neq 0=\int_0^1\int_0^1 f(x,y)\,dx\,dy.$$

This does not contradict Fubini's theorem, since

$$\int_0^1\int_0^1 |f(x,y)|\,dy\,dx=\infty.$$

(5) Let

$$\begin{aligned}Q&=\{(x,y)\in\mathbb{R}^2: x,y\geqslant 0\},\\ R&=\{(x,y)\in Q: x-1\leqslant y\leqslant x\},\\ S&=\{(x,y)\in Q: x-2\leqslant y\leqslant x-1\}\end{aligned}$$

and $f=\chi_S-\chi_R$. Since the two-dimensional Lebesgue measure of R and S is infinite, $f\notin L^1(\mathbb{R}^2)$. Define

$$g(x)=\int_{-\infty}^{\infty} f(x,y)\,dy=\begin{cases}-x, & 0\leqslant x\leqslant 1,\\ x-2, & 1\leqslant x\leqslant 2,\\ 0, & x\geqslant 2.\end{cases}$$

Then $\int_{-\infty}^{\infty} g(x)\,\mathrm{d}x = -1$. Similarly

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = 0$$

for every y, but

$$\int_{-\infty}^{\infty} h(y)\,dy = 0 \neq -1 = \int_{-\infty}^{\infty} g(x)\,dx.$$

THE END

Bibliography

[1] V.I. Bogachev, Measure Theory 1 and 2, Springer 2007.

[2] A.M. Bruckner, J.B. Bruckner, and B.S. Thomson, Real Analysis, Prentice-Hall, 1997.

[3] D.L. Cohn, Measure theory (2nd edition), Birkhäuser, 2013.

[4] L.C. Evans and R.F. Gariepy, Measure Theory and Fine Properties of Functions, CRC Press, 1992.

[5] G.B. Folland, Real Analysis. Modern Techniques and Their Applications (2nd edition), John Wiley & Sons, 1999.

[6] T. Jech, Set theory (3rd edition), Springer, 2003.

[7] F. Jones, Lebesgue Integration on Euclidean Space (revised edition), Jones and Bartlett Publishers, 2001.

[8] K.L. Kuttler, Modern Real Analysis, CRC Press, 1998.

[9] W. Rudin, Real and Complex Analysis, McGraw-Hill, 1986.

[10] D.A. Salamon, Measure and Integration, EMS, 2016,

[11] E. Stein and R. Sakarchi, Real Analysis: Measure Theory, Integration, and Hilbert Spaces, Princeton University Press, 2005.

[12] T. Tao, Introduction to Measure Theory, American Mathematical Society, 2011.

[13] M.E. Taylor, Measure theory and integration, American Mathematical Society, 2006.

[14] R.L. Wheeden and A. Zygmund, Measure and Integral: An introduction to Real Analysis, Marcel Dekker, 1977.

[15] W.P. Ziemer, Modern Real Analysis, PWS Publishing Company, 1995.

[16] J. Yeh, Real Analysis, Theory of Measure and Integration (2nd edition), World Scientific, 2006.