Measure and Integral
1 MEASURE THEORY ........................................... 1
  1.1 Outer measures ....................................... 1
  1.2 Measurable sets .................................... 5
  1.3 Measures ............................................. 10
  1.4 The distance function ................................ 14
  1.5 Approximations of measurable sets ................. 17
  1.6 Metric outer measures ................................ 27
  1.7 Lebesgue measure ................................... 31
  1.8 Invariance properties of the Lebesgue measure .... 37
  1.9 Lebesgue measurable sets ............................ 38
  1.10 A nonmeasurable set ................................ 43
  1.11 The Cantor set ..................................... 46

2 MEASURABLE FUNCTIONS .................................... 49
  2.1 Calculus with infinities .............................. 49
  2.2 Measurable functions ................................. 50
  2.3 Cantor-Lebesgue function ............................ 55
  2.4 Lipschitz mappings on \( \mathbb{R}^n \) ..................... 57
  2.5 Limits of measurable functions ...................... 60
  2.6 Almost everywhere .................................. 61
  2.7 Approximation by simple functions ................. 63
  2.8 Modes of convergence ................................ 66
  2.9 Egoroff’s and Lusin’s theorems ...................... 70

3 INTEGRATION .................................................. 76
  3.1 Integral of a nonnegative simple function .......... 76
  3.2 Integral of a nonnegative measurable function .... 78
  3.3 Monotone convergence theorem ........................ 81
  3.4 Fatou’s lemma ........................................ 85
  3.5 Integral of a signed function .......................... 86
  3.6 Dominated convergence theorem ...................... 90
  3.7 Lebesgue integral .................................... 95
  3.8 Cavalieri’s principle .................................. 106
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9</td>
<td>Lebesgue and Riemann</td>
<td>110</td>
</tr>
<tr>
<td>3.10</td>
<td>Fubini’s theorem</td>
<td>114</td>
</tr>
<tr>
<td>3.11</td>
<td>Fubini’s theorem for Lebesgue measure</td>
<td>119</td>
</tr>
</tbody>
</table>
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 open. Existence of a nonmeasurable set for the Lebesgue outer measure is shown by the axiom of choice.

1 Measure theory

1.1 Outer measures

Let $X$ be a set and consider a mapping on the collection of subsets of $X$.

**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) $\mu^*(A) \leq \mu^*(B)$ whenever $A \subset B \subset X$ and
3. (countable subadditivity) $\mu^* (\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i).

Countable subadditivity implies finite subadditivity, since we can always add an infinite number of empty sets.

**The Moral:** An outer measure is a general tool to measure size of a set. It is easy to construct an outer measure and all subsets of $X$ can be measured by an outer measure, but countable subadditivity alone does not produce a proper measure theory.

**Warning:** It may happen that the equality $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ fails when $A \cap B = \emptyset$. This means that an outer measure is not necessarily additive on pairwise disjoint sets.

**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 $\mu^*$ is an outer measure on $X$. However, if $A = \{1\}$ and $B = \{2\}$, then

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

Observe that inequality $\leq$ holds by subadditivity.
Examples 1.3:

(1) (The discrete measure) Let
\[ \mu^*(A) = \begin{cases} 
1, & A \neq \emptyset, \\
0, & A = \emptyset. 
\end{cases} \]

(2) (The Dirac measure) Let \( x_0 \in X \) be a fixed point and define
\[ \mu^*(A) = \begin{cases} 
1, & x_0 \in A, \\
0, & x_0 \notin A. 
\end{cases} \]

This is called the Dirac measure at \( x_0 \).

(3) (The counting measure) Let \( \mu^*(A) \) be the (possibly infinite) number of points in \( A \).

(4) (The Lebesgue measure) Let \( X = \mathbb{R}^n \) and consider the \( n \)-dimensional interval
\[ I = \{ x \in \mathbb{R}^n : a_i \leq x_i \leq b_i, i = 1, \ldots, n \} = [a_1, b_1] \times \cdots \times [a_n, b_n] \]
with sides parallel to the coordinate axes. The geometric volume of \( I \) is
\[ \text{vol}(I) = (b_1 - a_1)(b_2 - a_2) \cdots (b_n - a_n). \]

The Lebesgue outer measure of a set \( A \subset \mathbb{R}^n \) is defined as
\[ m^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \text{vol}(I_i) : A \subset \bigcup_{i=1}^{\infty} I_i \right\}. \]
By the definition of infimum, this means that for every $\epsilon > 0$ there are intervals $I_i$, $i = 1, 2, \ldots$, such that $A \subset \bigcup_{i=1}^{\infty} I_i$ and

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

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

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{covering_by_intervals.png}
\caption{Covering by intervals.}
\end{figure}

**Claim:** $m^*$ is an outer measure

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

$$0 \leq m^*(\emptyset) \leq \text{vol}([-\epsilon^{1/n}/2, \epsilon^{1/n}/2]^n) = \epsilon.$$  

By letting $\epsilon \to 0$, 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.

(2) We may assume that $m^*(B) < \infty$, for otherwise the claim is clear. For every $\epsilon > 0$ there are intervals $I_i$, $i = 1, 2, \ldots$, such that $B \subset \bigcup_{i=1}^{\infty} I_i$ and

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

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

$$m^*(A) \leq \sum_{i=1}^{\infty} \text{vol}(I_i) < m^*(B) + \epsilon.$$
By letting \( \varepsilon \to 0 \), we conclude \( m^*(A) \leq m^*(B) \).

[3] We may assume that \( m^*(A_i) < \infty \) for every \( i = 1, 2, \ldots, \), for otherwise the claim is clear. 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

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

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

\[
m^*\left(\bigcup_{i=1}^\infty A_i\right) \leq \sum_{i,j=1}^\infty \text{vol}(I_{j,i}) = \sum_{i=1}^\infty \sum_{j=1}^\infty \text{vol}(I_{j,i}) \leq \sum_{i=1}^\infty \left( m^*(A_i) + \frac{\varepsilon}{2^i} \right) = \sum_{i=1}^\infty m^*(A_i) + \varepsilon.
\]

The claim follows by letting \( \varepsilon \to 0 \).

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

\[
\mathcal{H}^s_\delta(A) = \inf \left\{ \sum_{i=1}^\infty \text{diam}(B_i)^s : A \subset \bigcup_{i=1}^\infty B_i, \text{diam}(B_i) \leq \delta \right\}
\]

and

\[
\mathcal{H}^s(A) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(A) = \sup_{\delta > 0} \mathcal{H}^s_\delta(A).
\]

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

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

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

is an outer measure on \( X \). Moreover, if \( \rho \) is monotone and countably subadditive on \( \mathcal{F} \), then \( \mu^* = \rho \) on \( \mathcal{F} \). (Exercise)

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

- For every \( \delta > 0 \) there are \( A_i \in \mathcal{F} \), \( i = 1, 2, \ldots, \), such that \( X = \bigcup_{i=1}^\infty A_i \) and \( \text{diam}(A_i) \leq \delta \).

- For every \( \delta > 0 \) there is \( A \in \mathcal{F} \) such that \( \rho(A) \leq \delta \) and \( \text{diam}(A) \leq \delta \).

For \( 0 < \delta \leq \infty \) and \( A \subset X \), we define

\[
\mu^*_\delta(A) = \inf \left\{ \sum_{i=1}^\infty \rho(A_i) : A_i \in \mathcal{F}, A \subset \bigcup_{i=1}^\infty A_i, \text{diam}(A_i) \leq \delta \right\}.
\]
The first assumption guarantees that we can cover any set \( A \) with sets in \( \mathcal{F} \) and the second assumption implies \( \mu^*_\delta(\varnothing) = 0 \). It is not difficult to see that \( \mu^*_\delta \) is an outer measure (exercise), but it is usually not additive and not a Borel measure. Clearly,

\[
\mu^*_\delta(A) \leq \mu^*\delta(A) \quad \text{when} \quad 0 < \delta < \delta' \leq \infty.
\]

Thus we may define

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

The outer measure \( \mu^* \) has much better properties than \( \mu^*_\delta \). For example, it is always a Borel measure, see Theorem 1.41 and Remarks 1.42. Moreover, if the members of \( \mathcal{F} \) are Borel sets, then \( \mu^* \) is Borel regular, see Definition 1.27.

**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

The next definition is so-called Carathéodory criterion for measurability.

**Definition 1.4.** A set \( A \subset X \) on \( \mu^* \)-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

\[
\mu^*(E) \leq \mu^*(E \cap A) + \mu^*(E \setminus A).
\]

This means that the inequality \( \leq \) holds always.

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

\[
\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \setminus A) = \mu^*(A) + \mu^*(B \setminus A).
\]

This shows that an outer measure behaves additively on measurable subsets.
(3) If $\mu^*(A) = 0$, then $A$ is $\mu^*$-measurable. In other words, all sets of measure zero are measurable.

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

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

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 $\mu^*$-measurable. The claim that $\emptyset$ is $\mu^*$-measurable follows from fact that $\mu^*(\emptyset) = 0$. On the other hand, a set $A$ is $\mu^*$-measurable if and only if $X \setminus A$ is $\mu^*$-measurable. Hence $X = X \setminus \emptyset$ is $\mu^*$-measurable. This can also be proved directly from the definition, since

$$
\mu^*(E) = \mu^*(E \cap \emptyset) + \mu^*(E \setminus \emptyset) = 0 + \mu^*(E)
$$

and

$$
\mu^*(E) = \mu^*(E \cap X) + \mu^*(E \setminus X) = 0 + \mu^*(E)
$$

from which it follows that $\emptyset$ and $X$ are $\mu^*$-measurable.

(5) The only measurable sets for the discrete measure are $\emptyset$ and $X$.

(6) All sets are measurable for the Dirac measure.
Example 1.6. (Continuation of Example 1.2) Let $X = \{1, 2, 3\}$ and, define an outer measure $\mu^*$ 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 $\mu^*$-measurable. In the same way we can see that all sets consisting of two points are not $\mu^*$-measurable. In this case only $\mu^*$-measurable sets are $\emptyset$ and $X$.

**Lemma 1.7.** The collection $\mathcal{M}$ of $\mu^*$-measurable sets is a $\sigma$-algebra, that is,

1. $\emptyset \in \mathcal{M}$,
2. $A \in \mathcal{M}$ implies $A^c = X \setminus A \in \mathcal{M}$ and
3. $A_i \in \mathcal{M}$ for every $i = 1, 2, \ldots$ implies $\bigcup_{i=1}^\infty A_i \in \mathcal{M}$.

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

**Remark 1.8.** $A_i \in \mathcal{M}$ for every $i = 1, 2, \ldots$ implies also that $\bigcap_{i=1}^\infty A_i \in \mathcal{M}$. To see this, observe that by de Morgan’s law

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

Hence the collection of $\mu^*$-measurable sets is not only closed with respect to countable unions and complements, but also with respect to countable intersections.

**Proof.**

1. $\mu^*(\emptyset) = 0$ implies that $\emptyset \in \mathcal{M}$.
2. $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A) = \mu^*(E \cap (X \setminus A)) + \mu^*(E \cap (X \setminus A))$ for every $E \subset X$. This implies that $A^c \in \mathcal{M}$.
3. Let $E \subset X$.

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

$$
\mu^*(E) = \mu^*(E \cap A_1) + \mu^*(E \setminus A_1) \quad (A_1 \in \mathcal{M}, E \text{ test set})
$$

$$(A_2 \in \mathcal{M}, E \setminus A_1 \text{ as a test set})$$

$$= \mu^*((E \setminus A_1) \cap A_2) + \mu^*(E \setminus A_1 \setminus A_2) + \mu^*(E \cap A_1)$$

$$(E \setminus A_1 \setminus A_2 = E \setminus (A_1 \cup A_2))$$

$$\geq \mu^*((E \setminus A_1 \cap A_2) \cup (E \cap A_1)) + \mu^*(E \setminus (A_1 \cup A_2)) \quad \text{(subadditivity)}$$

$$= \mu^*(E \cap (A_1 \cup A_2)) + \mu^*(E \setminus (A_1 \cup A_2))$$

$$= \mu^*(E \cap (A_1 \cup A_2)) \lor (E \setminus (A_1 \cup A_2)) = E \cap (A_1 \cup A_2)$$

for every $E \subset X$. By iteration, the same result holds for finitely many sets: If $A_i \in \mathcal{M}, i = 1, 2, \ldots, k$, then $\bigcup_{i=1}^k A_i \in \mathcal{M}$. 
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 $B_k = \bigcup_{i=1}^{k} A_i$, $k = 1,2,\ldots$ Then $B_k \subset B_{k+1}$ and

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

Define

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

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

![Figure 1.4: Covering by disjoint sets.](image)

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

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

By choosing $E = X$, this implies finite additivity on disjoint measurable sets. Observe, that $\leq$ holds by subadditivity.
The claim is clear for \( k = 1 \). Assume that the claim holds with index \( k \). Then

\[
\mu^*(E \cap B_{k+1}) = \mu^*((E \cap B_{k+1}) \cap B_k) + \mu^*((E \cap B_{k+1}) \setminus B_k)
\]

\( (B_k \in \mathcal{M}, E \cap B_{k+1} \text{ as a test set}) \)

\[
= \mu^*(E \cap B_k) + \mu^*(E \cap A_{k+1}) \quad (A_i \text{ are pairwise disjoint})
\]

\[
= \sum_{i=1}^{k} \mu^*(E \cap A_i) + \mu^*(E \cap A_{k+1}) \quad (\text{induction assumption})
\]

\[
= \sum_{i=1}^{k+1} \mu^*(E \cap A_i).
\]

**Step 4:** We have

\[
\sum_{i=1}^{k} \mu^*(E \cap A_i) = \mu^*(E \cap B_k) \quad \text{ (Step 3)}
\]

This implies that

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

By subadditivity

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

Consequently

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

whenever \( A_i \in \mathcal{M}, i = 1, 2, \ldots \), are disjoint. 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 \( A_i, i = 1, 2, \ldots \), disjoint. Then

\[
\mu^*(E) = \mu^*(E \cap B_k) + \mu^*(E \setminus B_k) \quad \text{ (step 1, } B_k \in \mathcal{M})
\]

\[
\geq \sum_{i=1}^{k} \mu^*(E \cap A_i) + \mu^*(E \setminus A), \quad k = 1, 2, \ldots \quad \text{ (Step 3, } B_k \subset A)
\]

This implies that

\[
\mu^*(E) \geq \lim_{k \to \infty} \sum_{i=1}^{k} \mu^*(E \cap A_i) + \mu^*(E \setminus A)
\]

\[
= \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \setminus A)
\]

\[
= \mu^*(E \cap A) + \mu^*(E \setminus A). \quad \text{ (Step 4)}
\]

Note that Step 4 is not really needed here. We could have used countable subadditivity instead as an inequality is enough here. But we will need Step 4 in the proof of Theorem 1.9. \( \square \)
1.3 Measures

From the proof of the previous lemma we see that an outer measure is countably additive on measurable and disjoint sets. This is a very useful property. As we saw before, 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 1.9.** (Countable additivity) Assume that $A_i \in \mathcal{M}, i = 1, 2, \ldots$, are $\mu^*$-measurable and disjoint sets. 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 disjoint measurable sets.

**Figure 1.5:** Disjoint sets.

**Proof.** By the proof of the previous lemma (Lemma 1.7, Step 4)
\[
\mu^* \left( E \cap \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu^*(E \cap A_i)
\]
whenever $A_i \in \mathcal{M}, i = 1, 2, \ldots$, are disjoint and $E \subset X$. The claim follows by choosing $E = X$. □
Definition 1.10. Assume that $\mathcal{M}$ is $\sigma$-algebra in $X$. A mapping $\mu : \mathcal{M} \to [0, \infty]$ is a measure on a measure space $(X, \mathcal{M}, \mu)$, if

1. $\mu(\emptyset) = 0$ and
2. $\mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i)$ whenever $A_i \in \mathcal{M}$, $i = 1, 2, \ldots$, are disjoint.

The Moral: A measure is a countably additive set function on a $\sigma$-algebra. An outer measure is defined on all subsets, but a measure is defined only on sets in the $\sigma$-algebra.

Remarks 1.11:

1. A measure $\mu$ is monotone on $\mathcal{M}$, since $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$ for every $A, B \in \mathcal{M}$ with $A \subset B$. In the same way we can see that $\mu$ is countably subadditive on $\mathcal{M}$.

2. 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 $\sigma$ is a $\sigma$-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 $\{2, 3\} \in \mathcal{M}$, but $\mu^*(\{2, 3\}) = 0$, but $\{2\} \notin \mathcal{M}$. ■

A measure is said to be complete, if $B \in \mathcal{M}$, $\mu(B) = 0$ and $A \subset B$ implies $A \in \mathcal{M}$. By monotonicity $\mu(A) = 0$. In other words, a measure is complete, if every subset of a set of measure zero in the $\sigma$-algebra belongs to the $\sigma$-algebra. For example, the measure space $(\mathbb{R}^n, \mathcal{B}, \mu)$, where $\mathcal{B}$ denotes the Borel sets and $\mu$ the Lebesgue measure is not complete, see Definition 1.25 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 $\sigma$-algebra. (Exercise)

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

$$\mu^*(E) = \inf \{ \mu(A) : E \subset A, A \in \mathcal{M} \}.$$ 

Assume that the measure space $(X, \mathcal{M}, \mu)$ is $\sigma$-finite, see Definition 1.23. Then every set in $\mathcal{M}$ is $\mu^*$-measurable and $\mu = \mu^*$ on $\mathcal{M}$. This means that $\mu^*$ is an extension of $\mu$. If the measure space $(X, \mathcal{M}, \mu)$ is complete, then the class of $\mu^*$-measurable functions is precisely $\mathcal{M}$. (Exercise)

The Moral: If the measure is $\sigma$-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.
(4) It is possible to develop theory also for signed or even complex valued measures. Assume that $\mathcal{M}$ is $\sigma$-algebra in $X$. A mapping $\mu: \mathcal{M} \to [-\infty, \infty]$ 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 and
\[ \mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i). \]

Next we give some examples of measures.

**Examples 1.12:**

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

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

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

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

**Theorem 1.13.** Assume that $\mu^*$ is an outer measure on $X$ and that sets $A_i \subset X$, $i = 1, 2, \ldots$, are $\mu^*$-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
\[ \lim_{i \to \infty} \mu^*(A_i) = \mu^* \left( \bigcup_{i=1}^{\infty} A_i \right). \]

**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.

**Remarks 1.14:**

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
\[
\mu^*(A) = \begin{cases} 
0, & A = \emptyset, \\
1, & A \text{ finite,} \\
2, & A \text{ infinite.}
\end{cases}
\]
Figure 1.6: Monotone sequences of sets.

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

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

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

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

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

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

\[ \mu^* \left( \bigcup_{i=1}^{\infty} A_i \right) = \mu^* \left( A_1 \cup \bigcup_{i=1}^{\infty} (A_{i+1} \setminus A_i) \right) \]

\[ = \mu^* (A_1) + \sum_{i=1}^{\infty} \mu^* (A_{i+1} \setminus A_i) \quad \text{(disjointness and measurability)} \]

\[ = \mu^* (A_1) + \sum_{i=1}^{\infty} \left( \mu^* (A_{i+1}) - \mu^* (A_i) \right) \]

\[ = \lim_{k \to \infty} \left( \mu^* (A_1) + \sum_{i=1}^{k} \left( \mu^* (A_{i+1}) - \mu^* (A_i) \right) \right) \]

\[ = \lim_{i \to \infty} \mu^* (A_{i+1}) = \lim_{i \to \infty} \mu^* (A_i). \]

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 \( A_1 \setminus A_i \subset A_1 \setminus A_{i+1} \).

\[ \mu^* \left( \bigcup_{i=1}^{\infty} (A_1 \setminus A_i) \right) = \lim_{i \to \infty} \mu^* (A_1 \setminus A_i) \quad (1) \]

\[ = \lim_{i \to \infty} \left( \mu^* (A_1) - \mu^* (A_i) \right) \]

\[ (\mu^* (A_1) = \mu^* (A_1 \cap A_i) + \mu^* (A_1 \setminus A_i)) \]

\[ = \mu^* (A_1) - \lim_{i \to \infty} \mu^* (A_i). \]

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 1.15.** Let \( A \subset \mathbb{R}^n \) with \( A \neq \emptyset \). Then

\[ \text{dist}(x, A) = \inf \{|x - y| : y \in A\} \]

is the distance from \( x \) to \( A \).
Remarks 1.16:

1. The distance between the nonempty sets \( A, B \subset \mathbb{R}^n \) is
   \[
   \text{dist}(A, B) = \inf \{|x - y| : x \in A \text{ and } y \in B\}.
   \]

2. The diameter of the nonempty set \( A \subset \mathbb{R}^n \) is
   \[
   \text{diam}(A) = \sup \{|x - y| : x, y \in A\}.
   \]

Lemma 1.17. 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
   \[
   \text{dist}(x, A) = |x_0 - x|.
   \]

The moral: 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 \( y_i \in A, i = 1, 2, \ldots \) such that
   \[
   \lim_{i \to \infty} |x - y_i| = \text{dist}(x, A).
   \]

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

Lemma 1.18. Let \( A \subset \mathbb{R}^n \) with \( A \neq \emptyset \). Then \( |\text{dist}(x, A) - \text{dist}(y, A)| \leq |x - y| \) for every \( x, y \in \mathbb{R}^n \).

The moral: 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 \( |x - z| \leq |x - y| + |y - z| \) for every \( z \in A \). For every \( \varepsilon > 0 \) there exists \( z' \in A \) such that \( |y - z'| \leq \text{dist}(y, A) + \varepsilon \). Thus
   \[
   \text{dist}(x, A) \leq |x - z'| \leq |x - y| + \text{dist}(y, A) + \varepsilon,
   \]

which implies
   \[
   |\text{dist}(x, A) - \text{dist}(y, A)| \leq |x - y| + \varepsilon.
   \]

By switching the roles of \( x \) and \( y \), we obtain
   \[
   |\text{dist}(x, A) - \text{dist}(y, A)| \leq |x - y| + \varepsilon.
   \]

This holds for every \( \varepsilon > 0 \), so that
   \[
   |\text{dist}(x, A) - \text{dist}(y, A)| \leq |x - y|. \tag*{□}
   \]
Lemma 1.19. Let \( A \subset \mathbb{R}^n \) be an open set with \( \partial A \neq \emptyset \). Define
\[
A_i = \left\{ x \in A : \text{dist}(x, \partial A) > \frac{1}{i} \right\}, \quad i = 1, 2, \ldots
\]
Then the sets \( A_i \) are open, \( A_i \subset A_{i+1}, i = 1, 2, \ldots \), and that \( A = \bigcup_{i=1}^\infty A_i \).

**The Moral:** 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
\[
\left\{ x \in A : \text{dist}(x, \partial A) > \frac{1}{i} \right\} = \left( \frac{1}{i}, \infty \right)
\]
is an open set. It is immediate that \( A_i \subset A_{i+1}, i = 1, 2, \ldots \).

Since \( A_i \subset A \) for every \( i = 1, 2, \ldots \), 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 \( \text{dist}(x, \partial A) \geq \varepsilon \). 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 1.20. If \( A \subset \mathbb{R}^n \) is an open with \( \partial A \neq \emptyset \), then \( \text{dist}(K, \partial A) > 0 \) for every compact subset \( K \) of \( A \).

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

**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, \( K = \{(x, y) \in \mathbb{R}^2 : y \geq e^x\} \) is closed and \( K \subset A \). However, \( \text{dist}(K, A) = 0 \).

**Remark 1.21.** In addition, the distance function has the following properties:

1. \( x \in \overline{A} \) if and only if \( \text{dist}(x, A) = 0 \),
2. \( \emptyset \neq A \subset B \) implies \( \text{dist}(x, A) \geq \text{dist}(x, B) \),
3. \( \text{dist}(x, A) = \text{dist}(x, \overline{A}) \) for every \( x \in \mathbb{R}^n \) and
4. \( \overline{A} = \overline{B} \) if and only if \( \text{dist}(x, A) = \text{dist}(x, B) \) for every \( x \in \mathbb{R}^n \).

(Exercise)

**Remark 1.22.** 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) $0 \leq f(x) \leq 1$ 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

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

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

1.5 Approximations 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.

**Definition 1.23.** A set $A \subset \mathbb{R}^n$ is $\sigma$-finite with respect to $\mu^*$ if we can write $A = \bigcup_{i=1}^{\infty} A_i$, where $A_i$, $i = 1, 2, \ldots$, are $\mu^*$-measurable and $\mu^*(A_i) < \infty$ for every $i = 1, 2, \ldots$

**Example 1.24.** The Lebesgue outer measure $m^*$ is $\sigma$-finite.

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

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

We shall show later that all open sets are Lebesgue measurable, see Example 1.43.

**Definition 1.25.** The collection $\mathcal{B}$ of Borel sets is the smallest $\sigma$-algebra containing the open subsets of $\mathbb{R}^n$.

**Remarks 1.26:**

1. An intersection of $\sigma$-algebras is an $\sigma$-algebra and thus

$$\mathcal{B} = \bigcap \{ \mathcal{M} : \mathcal{M} \text{ is an } \sigma\text{-algebra containing the open sets subsets of } \mathbb{R}^n \}.$$  

2. Since any $\sigma$-algebra is closed with respect to complements, the collection $\mathcal{B}$ of Borel sets can be also defined as the smallest $\sigma$-algebra containing, for example, the closed subsets. Note that the collection of Borel sets does not only contain open and closed sets, but it also contains, for example, the $G_\delta$-sets which are countable intersections of open sets and $F_\sigma$-sets which are countable unions of closed sets.
(3) The best way to show that a given set is a Borel set is to prove that it belongs to a $\sigma$-algebra containing open (or closed) sets.

**Definition 1.27.** Let $\mu^*$ be an outer measure on $\mathbb{R}^n$.

1. $\mu^*$ is called a Borel outer measure, if all Borel sets are $\mu^*$-measurable.
2. A Borel outer measure $\mu^*$ 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. $\mu^*$ is a Radon outer measure, if $\mu^*$ is Borel regular and $\mu^*(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$.

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

**Remarks 1.28:**

1. In particular, all open and closed sets are measurable for a Borel measure. Thus the collection of measurable sets is relatively large.
2. In general, an outer measure $\mu^*$ is called regular, if for every set $A \subset \mathbb{R}^n$ there exists a $\mu^*$-measurable set $B$ such that $A \subset B$ and $\mu^*(A) = \mu^*(B)$.
3. The local finiteness condition $\mu^*(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$ is equivalent with the condition $m^*(B(x,r)) < \infty$ for every $x \in \mathbb{R}^n$ and $r > 0$.

**Examples 1.29:**

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

**Example 1.30.** The Lebesgue outer measure $m^*$ is Borel regular.

**Reason.** 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

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

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 measure, this implies

$$m^*(A) \leq m^*(B) \leq m^*(B_i) = m^*\left(\bigcup_{j=1}^\infty I_{j,i}\right) \leq \sum_{j=1}^\infty \text{vol}(I_{j,i}) < m^*(A) + \frac{1}{i}.$$
By letting \( i \to \infty \), we conclude \( m^*(A) = m^*(B) \). \( \blacksquare \)

**Example 1.31.** The Lebesgue outer measure satisfies the property \( m^*(K) < \infty \) for every compact set \( K \subset \mathbb{R}^n \).

*Reason.* Since \( K \) 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 \( m^* \), this implies

\[
m^*(K) \leq \text{vol}(I) = (b_1 - a_1) \cdots (b_n - a_n) < \infty.
\]

\( \blacksquare \)

**Theorem 1.32.** Let \( \mu^* \) be a Radon measure on \( \mathbb{R}^n \), \( \mu^*(\mathbb{R}^n) < \infty \) and \( A \subset \mathbb{R}^n \) a \( \mu^* \)-measurable set. Then 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 can be approximated by closed sets from inside and open sets from outside in the sense of measure.

**Proof.** **Step 1:** Let

\[
\mathcal{F} = \{ A \subset \mathbb{R}^n : A \ \mu^* \text{-measurable, for every } \varepsilon > 0 \text{ there exists a closed } F \subset A \text{ such that } \mu^*(A \setminus F) < \varepsilon \text{ and an open } G \supset A \text{ such that } \mu^*(G \setminus A) < \varepsilon \}
\]
be the collection of measurable sets that has the required approximation property.

**Strategy:** We show that $F$ is a $\sigma$-algebra that contains the Borel sets. Borel regularity takes care of the rest.

It is clear that $\emptyset \in F$ and that $A \in F$ implies $\mathbb{R}^n \setminus A \in F$. Let $A_i \in F$, $i = 1, 2, \ldots$. We show that $\bigcap_{i=1}^{\infty} A_i \in F$. Let $\varepsilon > 0$. Since $A_i \in F$, there exist a closed set $F_i$ and an open set $G_i$ such that $F_i \subset A_i \subset G_i$, $\mu^*(A_i \setminus F_i) < \frac{\varepsilon}{2i+1}$ and $\mu^*(G_i \setminus A_i) < \frac{\varepsilon}{2i+1}$ for every $i = 1, 2, \ldots$. Then

\[
\mu^*\left(\bigcap_{i=1}^{\infty} A_i \setminus \bigcap_{i=1}^{\infty} F_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i \setminus F_i) < \varepsilon \sum_{i=1}^{\infty} \frac{1}{2i+1} < \varepsilon.
\]

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 1.13 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 open sets, $\bigcap_{i=1}^{k} G_i$ is an open set, it will do as an approximation from outside. This shows that $F$ is a $\sigma$-algebra.

Then we show that $F$ 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

\[
A = \bigcap_{i=1}^{\infty} A_i,
\]

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

The sets $A_i$, $i = 1, 2, \ldots$, are open, because $x \mapsto \text{dist}(x, A)$ is continuous. Since $\mu^*(\mathbb{R}^n) < \infty$, Theorem 1.13 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.
Thus $\mathcal{F}$ is $\sigma$-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 $\sigma$-algebra with this property. This proves the claim for the Borel sets.

**Step 2:** Assume then $A$ is a general $\mu^*$-measurable set. By Borel regularity there exists a Borel set $B_1 \supset A$ such that $\mu^*(B_1) = \mu^*(A)$ and a Borel set $B_2$ such that $\mathbb{R}^n \setminus A \subset \mathbb{R}^n \setminus B_2$ 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 $F \subset B_2 \subset A \subset B_1 \subset G$ and

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

It follows that

$$
\mu^*(G \setminus A) \leq \mu^*(G \setminus B_1) + \mu^*(B_1 \setminus A) \quad \text{(subadditivity, even = holds)}
$$

$$
< \varepsilon + \mu^*(B_1) - \mu^*(A) \quad (\mu^*(B_1 \setminus A) = \mu^*(B_1) - \mu^*(A), \mu^*(A) < \infty)
$$

$$
= \varepsilon \quad (\mu^*(B_1) - \mu^*(A) = 0)
$$

and

$$
\mu^*(A \setminus F) \leq \mu^*(A \setminus B_2) + \mu^*(B_2 \setminus F)
$$

$$
\leq \mu\left((\mathbb{R}^n \setminus B_2) \setminus (\mathbb{R}^n \setminus A)\right) + \varepsilon
$$

$$
= \mu^*(\mathbb{R}^n \setminus B_2) - \mu^*(\mathbb{R}^n \setminus A) + \varepsilon \quad (\mathbb{R}^n \setminus A \text{ measurable, } \mu^*(\mathbb{R}^n \setminus A) < \infty)
$$

$$
= \varepsilon. \quad (\mu^*(\mathbb{R}^n \setminus B_2) - \mu^*(\mathbb{R}^n \setminus A) = 0, \mu^*(\mathbb{R}^n) < \infty)
$$

$\square$

**Figure 1.8:** A covering of a closed set by distance sets.
Definition 1.33. Let $\mu^*$ be an outer measure on $\mathbb{R}^n$ and $E$ an arbitrary subset of $\mathbb{R}^n$. Then the restriction of $\mu^*$ to $E$ is defined to be

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

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

**The Moral:** The restriction of a measure is essentially a technical notion.

Remarks 1.34:

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

Lemma 1.35. Let $\mu^*$ be a Borel regular outer measure on $\mathbb{R}^n$. Suppose that $E \subset \mathbb{R}^n$ is $\mu^*$-measurable and $\mu^*(E) < \infty$. Then $\mu^*(E)$ is a Radon measure.

Remarks 1.36:

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

**Reason.** The one-dimensional Hausdorff measure $\mathcal{H}^1$ is a Borel regular outer measure but not a Radon measure on $\mathbb{R}^2$, because $\mathcal{H}^1(B(0,1)) = \infty$ and the closed unit ball $B(0,1)$ is a compact subset of $\mathbb{R}^2$. We shall discuss more about this later.
(2) If $E$ is a Borel set, then $\mu^*(E)$ is Borel regular even if $\mu^*(E) = \infty$.

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

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

**Claim:** $\nu$ is Borel regular.

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

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

$$= \mu^*(E) + \mu^*(B_1 \setminus E) \quad (E \subset B_1)$$

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

Let $A \subset \mathbb{R}^n$. Since $\mu^*$ 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 as a union of two Borel sets $C$ is a Borel set.

$$(\mu^*|E)(C) = \mu^*(C \cap E) \leq \mu^*(B_1 \cap C) \quad (E \subset B_1)$$

$$= \mu^*(B_1 \cap B_2) \quad (B_1 \cap C = B_1 \cap (B_2 \cup (\mathbb{R}^n \setminus B_1))) = B_1 \cap B_2)$$

$$\leq \mu^*(B_1 \cap A)$$

$$= \mu^*((B_1 \cap A) \cap E) + \mu^*((B_1 \cap A) \setminus E) \quad (E \text{ is } \mu^* \text{-measurable})$$

$$\leq \mu^*(E \cap A) + \mu^*(B_1 \setminus E) \quad (\text{monotonicity})$$

$$= (\mu^*|E)(A). \quad (\mu^*(B_1 \setminus E) = 0)$$

On the other hand, $A \subset C$ implies $(\mu^*|E)(A) \leq (\mu^*|E)(C)$. Consequently $(\mu^*|E)(A) = (\mu^*|E)(C)$ and $\mu^*|E$ is Borel regular. \hfill $\square$

**Theorem 1.37.** Let $\mu^*$ be a Radon measure on $\mathbb{R}^n$. Then the following conditions are equivalent:

1. $A \subset \mathbb{R}^n$ is $\mu^*$-measurable,
2. for every $\varepsilon > 0$ there exists a closet 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 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 measurability in Definition 1.4 depends only on the outer measure.
Remark 1.38. The result holds without the assumption $\mu^*(\mathbb{R}^n) < \infty$. Compare to Theorem 1.32

Proof: (1) $\Rightarrow$ (2) Let $\nu_i = \mu^*[B(0, i)]$, where $B(0, i) = \{x \in \mathbb{R}^n : |x| < i\}$, $i = 1, 2, \ldots$

By Lemma 1.35, $\nu_i$ is a Radon measure and $\nu_i(\mathbb{R}^n) \leq \mu^*[B(0, i)] < \infty$, for every $i = 1, 2, \ldots$. Since $A$ is $\mu^*$-measurable, $A$ is also $\nu_i$-measurable.

By Theorem 1.32, there exists an open set $G_i \supset A$ such that

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

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

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

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

$$\mu^*(G \setminus A) = \mu^* \left( \left( \bigcup_{i=1}^{\infty} (G_i \cap B(0, i)) \right) \setminus A \right)$$

$$= \mu^* \left( \bigcup_{i=1}^{\infty} (G_i \setminus B(0, i)) \right)$$

$$\leq \sum_{i=1}^{\infty} \mu^* \left( G_i \setminus B(0, i) \right)$$

$$\leq \sum_{i=1}^{\infty} \nu_i(G_i \setminus A) \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2i+1} < \varepsilon.$$

Thus $G$ will do as an approximation from outside.

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

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

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

(2) $\Rightarrow$ (1) For every $i = 1, 2, \ldots$ 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,

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

for every $i = 1, 2, \ldots$. This implies that

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

and consequently $\mu^*(A \setminus F) = 0$. Observe that $A = F \cup (A \setminus F)$, where $F$ is a Borel set and hence $\mu^*$-measurable. On the other hand, $\mu^*(A \setminus F) = 0$ so that $A \setminus F$ is $\mu^*$-measurable. The set $A$ is $\mu^*$-measurable as a union of two measurable sets. □
Remark 1.39. We can see from the proof that, for a Radon measure, an arbitrary measurable set differs from a Borel set only by a set of measure zero. A set $A$ is $\mu^*$-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 $\mathcal{F}_\sigma$ set. On the other hand, a set $A$ is $\mu^*$-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 $\mathcal{G}_\delta$ set.

Corollary 1.40. Let $\mu^*$ be a Radon measure on $\mathbb{R}^n$. Then the following claims are true:

1. (Outer measure) for every set $A \subset \mathbb{R}^n$,
   $$\mu^*(A) = \inf\{\mu^*(G) : A \subset G, \text{ G open}\}.$$

2. (Inner measure) for every $\mu^*$-measurable set $A \subset \mathbb{R}^n$,
   $$\mu^*(A) = \sup\{\mu^*(K) : K \subset A, \text{ K compact}\}.$$

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: If $\mu^*(A) = \infty$, the claim is clear. Hence we may assume that $\mu^*(A) < \infty$. 

Figure 1.10: Approximation by open sets in measure.
Step 1: Assume that $A$ is a Borel set and let $\varepsilon > 0$. Since $\mu^*$ is a Borel measure, the set $A$ is $\mu^*$-measurable. By Theorem 1.37, there exists an open set $G \ni A$ such that $\mu^*(G \setminus A) < \varepsilon$. Moreover,

$$\mu^*(G) = \mu^*(G \cap A) + \mu^*(G \setminus A) \quad (A \text{ is } \mu^*\text{-measurable})$$

$$= \mu^*(A) + \mu^*(G \setminus A) < \mu^*(A) + \varepsilon. \quad (A \subset G)$$

This implies the claim.

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

$$\mu^*(A) = \inf \{ \mu^*(B) : B \subset G, \ G \text{ open} \} \quad \text{(Step 1)}$$

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

On the other hand, by monotonicity,

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

and, consequently, the equality holds.

Assume first that $\mu^*(A) < \infty$ and let $\varepsilon > 0$. By Theorem 1.37, there exists a closed set $F \subset A$ such that $\mu^*(A \setminus F) < \varepsilon$. Since $F$ is $\mu^*$-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 $\mu^*(A) = \infty$. Denote

$$B_i = \{ x \in \mathbb{R}^n : i - 1 < |x| < i \}, \quad i = 1, 2, \ldots$$

Then $A = \bigcup_{i=1}^{\infty} (A \cap B_i)$ and by Theorem 1.9

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

because the sets $A \cap B_i, i = 1, 2, \ldots$, are pairwise disjoint and $\mu^*$-measurable. Since $\mu^*$ is a Radon measure, $\mu^*(A \cap B_i) \leq \mu^*(B_i) < \infty$. 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

$$\lim_{k \to \infty} \mu^* \left( \bigcup_{i=1}^{k} F_i \right) = \mu^* \left( \bigcup_{i=1}^{\infty} F_i \right) \quad \text{(Theorem 1.13)}$$

$$= \sum_{i=1}^{\infty} \mu^*(F_i) \quad \text{\text{(}F_i \text{ disjoint \(F_i \subset A \cap B_i\), Theorem 1.9)}$$

$$\geq \sum_{i=1}^{\infty} \left( \mu^*(A \cap B_i) - \frac{1}{2^i} \right) = \infty.$$
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, \text{ \( F \) closed} \} = \infty.
\]

Finally we pass over to compact sets. Assume that \( F \) is closed. Then the sets \( F \cap B(0,i), i = 1, 2, \ldots \), are closed and bounded and hence compact. By Theorem 1.13,
\[
\mu^*(F) = \mu^* \left( \bigcup_{i=1}^{\infty}(F \cap B(0,i)) \right) = \lim_{i \to \infty} \mu^*(F \cap B(0,i))
\]
and consequently
\[
\sup\{\mu^*(K) : K \subset A, \text{ \( K \) compact} \} = \sup\{\mu^*(F) : F \subset A, \text{ \( F \) closed} \}.
\]

1.6 Metric outer measures

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

**Theorem 1.41.** Let \( \mu^* \) be an outer measure on \( \mathbb{R}^n \). If \( \mu^* \) is a metric outer measure, that is,
\[
\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)
\]
for every \( A, B \subset \mathbb{R}^n \) such that \( \text{dist}(A,B) > 0 \), then \( \mu^* \) is a Borel 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.

**Proof.** We shall show that every closed set \( F \subset \mathbb{R}^n \) is \( \mu^* \)-measurable. It is enough to show that
\[
\mu^*(E) \geq \mu^*(E \cap F) + \mu^*(E \setminus F)
\]
for every \( E \subset \mathbb{R}^n \). If \( \mu^*(E) = \infty \), the claim is clear. 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
\[
A_i = \left\{ x \in A : \text{dist}(x,F) \geq \frac{1}{i} \right\}, \quad i = 1, 2, \ldots,
\]
Clearly \( \text{dist}(A_i,F) > \frac{1}{i} \) for every \( i = 1, 2, \ldots \) and \( A = \bigcup_{i=1}^{\infty} A_i \).

**Reason.** Since \( G \) is open \( \text{dist}(x,F) > 0 \) for every \( x \in A \). Thus for every \( x \in A \) there exists \( i \) such that \( x \in A_i \), and therefore \( A \subset \bigcup_{i=1}^{\infty} A_i \). On the other hand, \( A_i \subset A \), \( i = 1, 2, \ldots \), implies \( \bigcup_{i=1}^{\infty} A_i \subset A \). Thus the equality holds.

**Claim:** \( \lim_{i \to \infty} \mu^*(A_i) = \mu^*(A) \).
Figure 1.11: Exhaustion by distance sets.

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

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

Then we prove the reverse inequality. Remember that $A = \bigcup_{i=1}^{\infty} A_i$. Let

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

Then $A = A_i \cup \bigcup_{j=1}^{\infty} B_j$ and by subadditivity

$$\mu^*(A) \leq \mu^*(A_i) + \sum_{j=1}^{\infty} \mu^*(B_j).$$

It follows that

$$\mu^*(A) \leq \lim_{i \to \infty} \mu^*(A_i) + \lim_{i \to \infty} \sum_{j=1}^{\infty} \mu^*(B_j),$$

where

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

provided

$$\sum_{j=1}^{\infty} \mu^*(B_j) < \infty.$$

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

$$\sum_{j=1}^{k} \mu^*(B_{2j}) = \mu^* \left( \bigcup_{j=1}^{k} B_{2j} \right) \leq \mu^*(E) < \infty.$$
CHAPTER 1. MEASURE THEORY

Figure 1.12: Exhaustion by separated distance sets.

and

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

These together imply

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

Thus

\[ \mu^*(A) \leq \lim_{i \to \infty} \mu^*(A_i) \]

and consequently

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

Finally

\[ \mu^*(E \cap F) + \mu^*(E \setminus F) = \mu^*(E \cap F) + \mu^*(A) \]

\[ = \lim_{i \to \infty} \left[ \mu^*(E \cap F) + \mu^*(A_i) \right] \quad \text{(above, } A = E \setminus F) \]

\[ = \lim_{i \to \infty} \mu^* \left( (E \cap F) \cup A_i \right) \quad \text{(dist}(A_i,F) > 0) \]

\[ \leq \mu^*(E). \quad \text{(monotonicity)} \]
Remarks 1.42:

(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 can be represented as a union of Borel set and a set of measure zero, see Remark 1.39.

(2) The Carathéodory construction in Example 1.3 (7) 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.

Example 1.43. The Lebesgue outer measure \( m^\ast \) is a metric outer measure.

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

\[
\sum_{i=1}^{\infty} \text{vol}(I_i) < m^\ast(A \cup B) + \varepsilon.
\]

By subdividing each \( I_i \) into smaller intervals, we may assume that \( \text{diam}(I_i) < \text{dist}(A, B) \) for every \( i = 1, 2, \ldots \). 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 \( I''_i, i = 1, 2, \ldots \), 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

\[
m^\ast(A) + m^\ast(B) \leq \sum_{i=1}^{\infty} \text{vol}(I'_i) + \sum_{i=1}^{\infty} \text{vol}(I''_i) \leq \sum_{i=1}^{\infty} \text{vol}(I_i) < m^\ast(A \cup B) + \varepsilon.
\]

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

**THE MORAL:** This implies that the Lebesgue outer measure \( m^\ast \) is a Borel measure. Thus all Borel sets are \( m^\ast \)-measurable, in particular closed and open sets, are Lebesgue measurable. This means that the very useful properties of an outer measure on measurable sets are available for a large class of sets. By Examples 1.30 and 1.31 we can conclude that \( m^\ast \) is a Radon measure. We shall discuss this in more detail in the next section.
1.7 Lebesgue measure

We have already discussed the definition of the Lebesgue outer measure in Example 1.3. The main idea in calculating the $n$-dimensional volume or measure of a subset of $\mathbb{R}^n$ is to approximate the set by unions of other sets, whose geometry is simple and whose geometric volumes are known. Even though we talk about the volume in $\mathbb{R}^n$, we have area in $\mathbb{R}^2$ and length in $\mathbb{R}$.

We recall the definition of the Lebesgue outer measure. Let $X = \mathbb{R}^n$ and consider the $n$-dimensional interval

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

with $a_i, b_i \in \mathbb{R}$, $i = 1, 2, \ldots$. We allow intervals to be degenerate, that is, $b_i = a_i$ for some $i$. The volume of $I$ is

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

Remark 1.44. A general $n$-dimensional interval is $I = I_1 \times \cdots \times I_n$, where each $I_i$ is a one dimensional interval. Recall, that one dimensional intervals are of the form $[a, b], (a, b], [a, b)$ or $(a, b)$, where the endpoints can be $-\infty$ or $\infty$. The concept of a volume is defined to all intervals in a natural way. We have chosen closed and bounded intervals as a starting point, but other intervals would do as well.
Definition 1.45. The Lebesgue outer measure of an arbitrary set $A \subset \mathbb{R}^n$ is defined as

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

The Morals: The Lebesgue outer measure attempts to describe the volume of an arbitrary set by approximating it from outside by countable unions of closed intervals and their volumes.

Several questions arise:

1. How do we compute the Lebesgue measure of a given set?
2. Does the geometric measure of a nice set equal to the Lebesgue measure of the set?
3. Which properties of the geometric notion of a volume pass over to the Lebesgue measure?
4. What does it mean for a set to be Lebesgue measurable?

Remark 1.46. A set $A \subset \mathbb{R}^n$ is of Lebesgue outer measure zero if and only if for every $\varepsilon > 0$ there are are intervals $I_i$, $i = 1, 2, \ldots$, such that $A \subset \bigcup_{i=1}^{\infty} I_i$ and

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

Figure 1.14: A set of Lebesgue measure zero.
Examples 1.47:

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

**Reason.** (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
\[
m^*(\{x\}) \leq \text{vol}(Q) = \varepsilon,
\]
which implies that \( m^*(\{x\}) = 0 \).

(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 is of Lebesgue measure zero.

**Reason.** Let \( A = \{x_1, x_2, \ldots, x_i \in \mathbb{R}^n \). We give two ways to prove the claim.

(1) Let \( \varepsilon > 0 \) and \( Q_i, i = 1, 2, \ldots, \) be a closed \( n \)-dimensional cube with center \( x_i \) and side length \( \left( \frac{\varepsilon}{2} \right)^{1/n} \). Then
\[
m^*(A) \leq \sum_{i=1}^{\infty} \text{vol}(Q_i) \leq \sum_{i=1}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon,
\]
which implies that \( m^*(A) = 0 \).

(2) By subadditivity
\[
m^*(A) = m^* \left( \bigcup_{i=1}^{\infty} \{x_i\} \right) \leq \sum_{i=1}^{\infty} m^*(\{x_i\}) = 0.
\]

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

Remarks 1.48:

(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
\[
m^* J(A) = \inf \left\{ \sum_{i=1}^{k} \text{vol}(I_i) : A \subset \bigcup_{i=1}^{k} I_i, \ k = 1, 2, \ldots \right\},
\]
where \( A \subset \mathbb{R}^n \) is a bounded set. The Jordan outer measure will not be an outer measure, since it is only finitely subadditive instead of countably subadditive. It has the property \( J^*(A) = J^*(A) \) for every bounded \( A \subset \mathbb{R}^n \).

We can define the corresponding Jordan inner measure by
\[
m_* J(A) = \sup \left\{ \sum_{i=1}^{k} \text{vol}(I_i) : A \supset \bigcup_{i=1}^{k} I_i, \ k = 1, 2, \ldots \right\}
\]
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 \( \overline{A} \) is zero.

For example,
\[
m^* J(Q \cap [0,1]) = 1 \quad \text{and} \quad m_* J(Q \cap [0,1]) = 0,
\]
while \( m^*(Q \cap [0,1]) = 0 \), since \( Q \cap [0,1] \) 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.

(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 lengths are equal, that is \( b_1 - a_1 = \cdots = b_n - a_n \). Even balls will do, but this is more subtle (exercise).

(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
\[
(2j)^n = \text{vol}(Q_j) \leq \sum_{i=1}^{\infty} \text{vol}(I_i).
\]
Letting \( j \to \infty \), we see that \( \sum_{i=1}^{\infty} \text{vol}(I_i) = \infty \) 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 \( \text{diam}(Q) = \sqrt{n}l(Q) \), where \( l(Q) \) is the side length of \( Q \). From this we conclude that \( l(Q) = r/\sqrt{n} \) and thus
\[
Q = \left[ x_1 - \frac{r}{4\sqrt{n}}, x_1 + \frac{r}{4\sqrt{n}} \right] \times \cdots \times \left[ x_n - \frac{r}{4\sqrt{n}}, x_n + \frac{r}{4\sqrt{n}} \right].
\]
This implies
\[ m^*(G) \geq m^*(Q) = \left( \frac{r}{2\sqrt{n}} \right)^n > 0. \]

\[ \square \]

**Figure 1.15:** A cube inside a ball.

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

**Example 1.49.** The Lebesgue outer measure of a closed interval \( I \subset \mathbb{R}^n \) is equal to the volume of the interval, that is \( m^*(I) = \text{vol}(I) \). This means that the definition is consistent for closed intervals.

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

\[ \sum_{j=1}^{\infty} \text{vol}(I_j) < m^*(I) + \epsilon. \]

For every \( I_j \) there is an open interval \( J_j \) such that \( I_j \subset J_j \) and

\[ \text{vol}(J_j) \leq \text{vol}(I_j) + \frac{\epsilon}{2^j}. \]

This implies
\[ \sum_{j=1}^{\infty} \text{vol}(J_j) \leq \sum_{j=1}^{\infty} \left( \text{vol}(I_j) + \frac{\epsilon}{2^j} \right) = \sum_{j=1}^{\infty} \text{vol}(I_j) + \sum_{j=1}^{\infty} \frac{\epsilon}{2^j} = \sum_{j=1}^{\infty} \text{vol}(I_j) + \epsilon. \]
The collection of intervals $J_j, j = 1, 2, \ldots$, is an open covering of the compact set $I$. It follows that there is a finite subcovering $J_j, j = 1, 2, \ldots, k$.

**Claim:** $\text{vol}(I) \leq \sum_{j=1}^{k} \text{vol}(J_j)$.

We split $I$ into finitely many subintervals $K_1, \ldots, K_m$ such that

$$I = \bigcup_{i=1}^{m} K_i, \quad \text{vol}(I) = \sum_{i=1}^{m} \text{vol}(K_i),$$

and $K_i \subset J_{j_i}$ for some $j_i, j_i \leq k$. Note that more than one $K_i$ may belong to the same $J_{j_i}$. Then

$$\bigcup_{i=1}^{m} K_i \subset \bigcup_{j=1}^{k} J_j$$

and

$$\text{vol}(I) = \sum_{i=1}^{m} \text{vol}(K_i) \leq \sum_{j=1}^{k} \text{vol}(J_j)$$

$$\leq \sum_{j=1}^{\infty} \text{vol}(J_j) \leq \sum_{j=1}^{\infty} \text{vol}(I_j) + \epsilon \leq m^*(I) + 2\epsilon.$$

The claim follows by letting $\epsilon \to 0$. ■

![Figure 1.16: Lebesgue measure of an interval.](image)
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

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

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

$$m^*(A + x_0) = \inf \left\{ \sum_{i=1}^{\infty} \text{vol}(I_i + x_0) : A + x_0 \subset \bigcup_{i=1}^{\infty} (I_i + x_0) \right\}$$

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

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

$$m^*(E \cap (A + x_0)) + m^*(E \setminus (A + x_0))$$

$$= m^*((E - x_0) \cap A) + m^*((E - x_0) \setminus A) + m^*((E - x_0) \setminus (A + x_0))$$

$$= m^*((E - x_0) \cap A) + m^*((E - x_0) \setminus A) + m^*(E - x_0) \ (A \text{ is measurable})$$

$$= m^*(E) \ (\text{translation invariance})$$

for every $E \subset \mathbb{R}^n$. This shows that $A + x_0$ is Lebesgue measurable. The equivalence follows from this.

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 $A \subset \mathbb{R}^n$, $\delta > 0$ 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 [4] pages 65–80 or [12] 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 1.50.** 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 3.35.

### 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 measure, see the discussion after Example 1.43. In particular, all Borel sets are Lebesgue measurable.

**The Moral:** Open or closed sets are Lebesgue measurable and all sets obtained from these by countably many set theoretic operations, as complements
intersections and unions, will give a 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.

Remarks 1.51:

(1) By the continuity properties of an outer measure, see Theorem 1.13, it is possible to show that the Lebesgue measure of elementary sets, such as other than closed intervals and finite unions of arbitrary intervals, agrees with the notion of volume. For example, if \( I = (a_1, b_1) \times \cdots \times (a_n, b_n) \) is an open interval in \( \mathbb{R}^n \), then \( I \) can be written as a union of an increasing sequence of closed intervals \( I_i, i = 1, 2, \ldots \), and

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

This follows also from the fact that \( m^*(\partial I) = 0 \) (exercise).

(2) We say that the intervals \( I_i, i = 1, 2, \ldots \), are almost disjoint if their interiors are pairwise disjoint. Thus almost disjoint intervals may touch on the boundaries. If \( I_i, i = 1, 2, \ldots \), are almost disjoint, then

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

(3) Every nonempty open set in \( \mathbb{R}^n \) is a union of countably many almost disjoint dyadic intervals.

Reason. Define a closed dyadic cube to be a cube of the form

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

In order to get a uniform bound, we consider only dyadic cubes of side length at most one, that is, \( k \) is a nonnegative integer. Observe that the dyadic cubes for a fixed \( k \) are almost disjoint and cover \( \mathbb{R}^n \). The 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. Let us consider the collection \( \mathcal{D} \) of all dyadic cubes contained in the given open set. These cubes are not pairwise disjoint, but we may consider cubes in \( \mathcal{D} \) that are maximal in the sense of inclusion, which means that they are not contained any of the other cubes in \( \mathcal{D} \). This collection of cubes satisfies the required properties (exercise).

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

\[
\left( \frac{i_1}{2^k}, \frac{i_1+1}{2^k} \right) \times \cdots \times \left( \frac{i_n}{2^k}, \frac{i_n+1}{2^k} \right), \quad i_1, \ldots, i_n \in \mathbb{Z}, \quad k \in \mathbb{Z}.
\]
In the one dimensional case every open set is a union of countably many disjoint open intervals. The Lebesgue measure of an open set is the sum of volumes of these intervals.

**Theorem 1.52.** 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. If $m^*(A) < \infty$, for every $\varepsilon > 0$, there exists a finite union $B = \bigcup_{i=1}^k Q_i$ of closed cubes such that $m^*(A \triangle B) < \varepsilon$. Here
   \[ A \triangle B = (A \setminus B) \cup (B \setminus A) \]
   is the symmetric difference of sets $A$ and $B$.
5. $m^*(A) = \inf\{m^*(G) : A \subset G, \ G \text{ open}\}$.
6. $m^*(A) = \sup\{m^*(K) : K \subset A, \ K \text{ compact}\}$.

**Proof.**

(1) and (2) The claims follow from Theorem 1.37.

[3] Let $\varepsilon > 0$. Since $m^*(A) < \infty$, by Corollary 1.40, there is a compact set $K \subset A$ such that $m^*(K) > m^*(A) - \varepsilon$. This implies
   \[
   m^*(A \setminus K) = m^*(A) - m^*(K) < \varepsilon.
   \]
Take cubes $Q_i$, $i = 1, 2, \ldots$, such that $A \subset \bigcup_{i=1}^{\infty} Q_i$ and

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

Since $m^*(A) < \infty$ the series on the left-hand side converges and thus there exists $k$ such that $\sum_{i=k+1}^{\infty} \text{vol}(Q_i) < \frac{\varepsilon}{2}$. Let $B = \bigcup_{i=1}^{k} Q_i$. Then

$$m^*(A \Delta B) = m^*(A \setminus B) + m^*(B \setminus A) \leq m^* \left( \bigcup_{i=k+1}^{\infty} Q_i \right) + m^* \left( \bigcup_{i=1}^{k} Q_i \setminus A \right) \leq m^* \left( \bigcup_{i=k+1}^{\infty} Q_i \right) + m^* \left( \bigcup_{i=1}^{\infty} Q_i \right) - m^*(A) \leq \sum_{i=k+1}^{\infty} \text{vol}(Q_i) + \sum_{i=1}^{\infty} \text{vol}(Q_i) - m^*(A) < \varepsilon.$$

The claims follow from Corollary 1.40. \hfill \Box

**The Moral:** The Jordan content of a set is defined by approximating the set from the inside and outside by finite unions of intervals, see Remark 1.48 (2). The Lebesgue measure is given by a two step approximation process. First we approximate the set from the outside by open sets and from the inside by compact sets and then we approximate the open sets from inside and compact sets from...
outside by finite unions of intervals. The Lebesgue measurable sets are precisely those sets for which these approximations coincide. If we allow epsilon of error both from inside and outside, we can approximate by a finite union of cubes.

By the previous approximation properties we obtain a characterization of Lebesgue measurable sets. Recall that a \( \mathcal{G}_\delta \) set is a countable intersection of open sets and a \( \mathcal{F}_\sigma \) set is a countable union of closed sets. In particular, \( \mathcal{G}_\delta \) and \( \mathcal{F}_\sigma \) sets are Borel sets. The following claims are equivalent, see Remark 1.39:

1. \( A \) is Lebesgue measurable.
2. \( A \) is a \( \mathcal{G}_\delta \) set with a set of measure zero removed.
3. \( A \) is a union of a \( \mathcal{F}_\sigma \) set and a set of measure zero.
4. \( \inf \{ m^*(G) : A \subset G, \text{G open} \} = \sup \{ m^*(K) : K \subset A, \text{K compact} \} \).

**The Moral:** An arbitrary Lebesgue measurable set differs from a Borel set only by a set of measure zero.

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 complete.

**Remark 1.53.** Let \( n, m \geq 1 \) be natural numbers.

1. If \( A \subset \mathbb{R}^n \) and \( B \subset \mathbb{R}^m \), then \( (m^{n+m})^*(A \times B) \leq (m^n)^*(A)(m^m)^*(B) \).
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 1.54.** The Lebesgue measure is defined to be the Lebesgue outer measure on the \( \sigma \)-algebra of Lebesgue measurable sets. We denote the Lebesgue measure by \( \mu \). In particular, the Lebesgue measure is countably additive on pairwise disjoint Lebesgue measurable sets.

**Remark 1.55.** 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 \( A_i \subset \mathbb{R}^n, i = 1, 2, \ldots \), 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) \( \mu((0,1)^n) = 1 \).
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 $2^{-k}$, $k = 1, 2, \ldots$. By additivity and translation invariance

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

A similar argument can be done also when $k = 0, -1, -2, \ldots$ 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 $G \subset \mathbb{R}^n$. The claim follows, since $m(A) = \inf\{m(G) : A \subset G, G \text{ open}\}$ for every $A \subset \mathbb{R}^n$.

1.10 A nonmeasurable set

It is natural to ask whether Lebesgue outer measure is countably additive on disjoint sets. This is true when all sets involved are Lebesgue measurable, but, as we shall see, additivity can break down if the sets are nonmeasurable. 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 measure. Vitali was the first to show that such a set exists using the axiom of choice. It has turned out later, that this cannot be done without the axiom of choice.

**Theorem 1.56.** 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

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

**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 the 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 $E_x$ contains all rational numbers. 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. To see this, suppose for contradiction that $E$ is Lebesgue measurable. Then the translated sets $E + q = \{x + q : x \in E\}$, $q \in \mathbb{Q}$, 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 disjoint and Lebesgue measurable, by countable additivity and translation invariance,

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

which is $0$ if $m^*(E) = 0$ and $\infty$ if $m^*(E) > 0$. On the other hand,

$$0 < 1 = m^*([0, 1]) \leq m^*(\bigcup_{q \in [-1, 1] \cap \mathbb{Q}} (E + q)) \leq m^*([-1, 2]) = 3 < \infty.$$  

This is a contradiction and thus $E$ cannot be Lebesgue measurable. □
Remark 1.57. 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

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

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,

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

This is a contradiction and thus $B$ cannot be Lebesgue measurable. \[\blacksquare\]

Remarks 1.58:

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

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

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 $\mathbb{R}^n$. This implies that the projection $\pi : \mathbb{R}^n \rightarrow \mathbb{R}, \pi(x) = x_1$ does not map measurable sets with the $n$-dimensional Lebesgue measure to Lebesgue measurable sets with respect to one-dimensional Lebesgue measure. We shall see more examples later.
Remark 1.59. 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 requires use of the axiom of choice. In particular, the pieces cannot be Lebesgue measurable, because otherwise the additivity fails.

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 \((\frac{1}{3}, \frac{2}{3})\). Thus, \( I_{1,1} \) is the open middle third of \( C_0 \). At the second steps 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 \( k \)th step, we obtain \( 2^{k-1} \) open pairwise disjoint intervals \( I_{k,i}, i = 1, \ldots, 2^{k-1} \), and denote

\[
C_0 = [0, 1], \quad C_k = C_{k-1} \setminus \bigcup_{i=1}^{2^{k-1}} I_{k,i}, \quad k = 1, 2, 3, \ldots
\]

Note:

\[
C_k = \bigcup_{a_1, \ldots, a_k \in \{0, 2\}} \left\{ \sum_{i=1}^{k} a_i 3^{-i} : a_i \in \{0, 2\}, i = 1, 2, \ldots \right\}.
\]

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

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

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

Note:

\[
C = \left\{ \sum_{i=1}^{\infty} a_i 3^{-i} : a_i \in \{0, 2\}, i = 1, 2, \ldots \right\}.
\]

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

Claim: \( C \) is uncountable.

Reason. Suppose that \( C = \{x_1, x_2, \ldots \} \) is countable. Let \( I_1 \) be an interval in the construction of the Cantor set such that \( x_1 \notin I_1 \). Then let \( I_2 \) be an interval in the construction of the Cantor set such that \( x_2 \notin I_2 \) and \( I_2 \subset I_1 \). By continuing this way, we obtain a decreasing sequence of closed intervals \( I_i \subset I_{i+1}, i = 1, 2, \ldots \), such that \( C \cap \bigcap_{i=1}^{\infty} I_i = \emptyset \). On the other hand, \( \bigcap_{i=1}^{\infty} I_i \neq \emptyset \) and \( \bigcap_{i=1}^{\infty} I_i \subset C \), which implies \( C \cap \bigcap_{i=1}^{\infty} I_i \neq \emptyset \). This is a contradiction. \( \blacksquare \)
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. Note that $C$ contains much more points than the end points \( \{\frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{2}{9}, \frac{7}{9}, \frac{8}{9}, \frac{1}{27}, \ldots\} \) 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).

We show that $m^*(C) = 0$. This means that $C$ is an uncountable set of measure zero.

**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 1.13 we have

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

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 $(2/3)^k$. This is arbitrarily small by choosing $k$ large enough. 

\[\]
CHAPTER 1. MEASURE THEORY

Remark 1.60. 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 \( i = 1,2,\ldots \) We denote this as \( x = .\alpha_1\alpha_2\ldots \). 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 \( \alpha_i \)'s are nonzero. For example

\[
\frac{4}{9} = .11000\cdots = .10222\cdots
\]

The reason for this is that

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

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 \( x = .1\alpha_2\alpha_3\ldots \). If \( x \in [0,1] \setminus I_{1,1} \), then \( x = .0\alpha_2\alpha_3\ldots \) or \( x = .2\alpha_2\alpha_3\ldots \). In either case the value of \( \alpha_1 \) 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 \( k \)th, stage of the construction we remove \( 2^{k-1} \) centrally situated open intervals each of length \( l_k \), \( k = 1,2,\ldots \), with \( l_1 + 2l_2 + \cdots + 2^{k-1}l_k < 1 \). If \( l_k \), \( k = 1,2,\ldots \), are chosen small enough, then

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

In this case, we have

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

and \( C \) is called a fat Cantor set. Note that \( C \) is a nowhere dense and perfect set of positive Lebesgue measure (exercise).
The class of measurable functions will play a central role in
the integration theory. This class is closed under usual op-
erations and limits, but certain unexpected features occur.
Measurable functions can be approximated by simple func-
tions. 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 ±∞. 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 define \( a + \infty = \infty + a = \infty \) for every \( a \in [0, \infty] \) and

\[
a \cdot \infty = \infty \cdot a = \begin{cases} 
\infty, & a > 0, \\
0, & a = 0.
\end{cases}
\]

With these definitions the standard commutative, associative and distributive
rules hold in [0, ∞] in the usual manner. Cancellation properties have to be treated
with some care. For example, \( a + b = a + c \) implies \( b = c \) only when \( a < \infty \) and
\( ab = ac \) implies \( b = c \) only when \( 0 < a < \infty \). This is related to the fact that \( \infty - \infty \)
and \( \infty / \infty \) are indeterminate in the sense that we cannot assign value to them without
breaking the rules of algebra. A general rule of thumb 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 \( x_i \in [0, \infty], 
\)
\( i = 1, 2, \ldots \), 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 is a measure on measurable sets.

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

\[
\{ x \in X : f(x) > a \}
\]

is \( \mu \)-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

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

is a Lebesgue measurable function.

   **Reason.**

   \[
   \{ x \in \mathbb{R}^n : f(x) > a \} = \begin{cases} \mathbb{R}^n, & a < 0, \\ A, & 0 \leq a < 1, \\ \varnothing, & a \geq 1. \end{cases}
   \]

   This implies that there exists a nonmeasurable function with respect to the Lebesgue measure.

3. The previous remark hold 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 Definition 2.27.

4. If all sets are measurable, then all functions are measurable.

5. If the only measurable sets are \( \emptyset \) and \( X \), then only constant functions are measurable.
Remark 2.3. For \( \mu \)-measurable subsets \( A \subset X \) and a function \( f : A \to [\infty, \infty] \), we consider the zero extension \( \tilde{f} : X \to [\infty, \infty] \),

\[
\tilde{f}(x) = \begin{cases} 
  f(x), & x \in A, \\
  0, & x \in X \setminus A. 
\end{cases}
\]

Then \( f \) is \( \mu \)-measurable if and only if \( \tilde{f} \) is \( \mu \)-measurable.

**The Moral:** A function defined on a subset is measurable if and only if its zero extension is measurable. This allows us to consider functions defined on subsets.

**Lemma 2.4.** Let \( \mu \) be a measure on \( X \) and \( f : X \to [\infty, \infty] \). Then the following claims are equivalent:

1. \( f \) is \( \mu \)-measurable,
2. \( \{x \in X : f(x) \geq a\} \) is \( \mu \)-measurable for every \( a \in \mathbb{R} \),
3. \( \{x \in X : f(x) < a\} \) is \( \mu \)-measurable for every \( a \in \mathbb{R} \),
4. \( \{x \in X : f(x) \leq a\} \) is \( \mu \)-measurable for every \( a \in \mathbb{R} \).

**Proof.** The equivalence follows from the fact that the collection of \( \mu \)-measurable sets is a \( \sigma \)-algebra, see Lemma 1.7.

\[
\begin{align*}
(1) \Rightarrow (2) \quad & \{x \in X : f(x) > a\} = \bigcap_{i=1}^{\infty} \{x \in X : f(x) > a - \frac{1}{i}\}. \\
(2) \Rightarrow (3) \quad & \{x \in X : f(x) < a\} = X \setminus \{x \in X : f(x) \geq a\}. \\
(3) \Rightarrow (4) \quad & \{x \in X : f(x) \leq a\} = \bigcap_{i=1}^{\infty} \{x \in X : f(x) < a + \frac{1}{i}\}. \\
(4) \Rightarrow (1) \quad & \{x \in X : f(x) < a\} = X \setminus \{x \in X : f(x) > a\}. \qed
\end{align*}
\]

**Lemma 2.5.** A function \( f : X \to [\infty, \infty] \) is \( \mu \)-measurable if and only if \( f^{-1}((-\infty)) \) and \( f^{-1}(\infty) \) are \( \mu \)-measurable and \( f^{-1}(B) \) is \( \mu \)-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 \( \mu \)-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.** [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 \( \mu \)-measurable.

Let

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

**Claim:** \( \mathcal{F} \) is a \( \sigma \)-algebra.
Reason. Clearly \( \emptyset \in \mathcal{F} \). If \( B \in \mathcal{F} \), then \( f^{-1}(\mathbb{R} \setminus B) = X \setminus f^{-1}(B) \) is \( \mu \)-measurable and thus \( \mathbb{R} \setminus B \in \mathcal{F} \). If \( B_i \in \mathcal{F}, \ i = 1, 2, \ldots \), then

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

is \( \mu \)-measurable and thus \( \bigcup_{i=1}^{\infty} B_i \in \mathcal{F} \). \( \blacksquare \)

Then we show that \( \mathcal{F} \) contains all open subsets of \( \mathbb{R} \). Since every open set in \( \mathbb{R} \) is a countable union of pairwise disjoint open intervals and \( \mathcal{F} \) is a \( \sigma \)-algebra, it is enough to show that every open interval \((a, b) \in \mathcal{F} \). 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 \( \mu \)-measurable. This implies that \( f^{-1}((a, b)) \) is \( \mu \)-measurable. Since \( \mathcal{F} \) is a \( \sigma \)-algebra that contains open sets, it also contains the Borel sets.

Choose \( B = (a, \infty) \) 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 \( \mu \)-measurable. \( \square \)

Remarks 2.7:

1. If \( f : X \to \mathbb{R} \) is measurable and \( g : \mathbb{R} \to \mathbb{R} \) is continuous, then the composed function \( g \circ f \) is measurable.

Reason. We shall show that the preimage \((g \circ f)^{-1}(G)\) of an open set \( G \subset \mathbb{R} \) is measurable. Note that \((g \circ f)^{-1}(G) = f^{-1}(g^{-1}(G))\), since

\[
x \in (g \circ f)^{-1}(G) \iff (g \circ f)(x) \in G \iff g(f(x)) \in G \iff f(x) \in g^{-1}(G) \iff x \in f^{-1}(g^{-1}(G)).
\]

Since \( g \) is continuous, the preimage \( g^{-1}(G) \) of an open set \( G \) is open. Since \( f \) is measurable, the preimage \( f^{-1}(g^{-1}(G)) \) on an open set \( g^{-1}(G) \) is measurable. Thus \((g \circ f)^{-1}(G)\) is a measurable set and \( g \circ f \) is a measurable function. \( \blacksquare \)

In fact, it is enough to assume that \( g \) is a Borel function, that is, the preimage of every Borel set is a Borel set. There is a version of this statement also for extended real valued functions. As we shall see in Section 2.3, a composed function of two measurable functions is not measurable, in general.

2. We shall briefly discuss an abstract version of a definition of a measurable function. 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 \( \sigma \)-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.
CHAPTER 2. MEASURABLE FUNCTIONS

(3) 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 the previous remark, then the composed function \(g \circ f\) is measurable. This follows directly from the abstract definition of measurability. We might be tempted to conclude that the composed function of Lebesgue measurable functions is Lebesgue measurable. Consider the case \(X = Y = Z = \mathbb{R}\) and \(\mathcal{P}\) is the Borel sets. Since \(g\) is Lebesgue measurable, \(\mathcal{M}\) contains Lebesgue measurable sets. If the composed function \(g \circ f\) were Lebesgue measurable, then the preimage of a Lebesgue measurable set should be Lebesgue measurable, that is, \(f^{-1}(A) \in \mathcal{M}\) whenever \(A \in \mathcal{N}\). However, this is not true. This means that we cannot replace Borel sets by measurable sets in Lemma 2.5. We shall give an example of this in Section 2.3.

**Lemma 2.8.** If \(f, g : X \to [-\infty, \infty]\) are \(\mu\)-measurable functions, then

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

is a \(\mu\)-measurable set.

*Proof:* Since the set of rational numbers is countable, we have \(Q = \bigcup_{i=1}^{\infty} \{q_i\}\). If \(f(x) > g(x)\), there exists \(q_i \in Q\) 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 \(\mu\)-measurable set. \(\square\)

**Remark 2.9.** The sets

\[
\{x \in X : f(x) \leq g(x)\} = X \setminus \{x \in X : f(x) > g(x)\}
\]

and

\[
\{x \in X : f(x) = g(x)\} = \{x \in X : f(x) \leq g(x)\} \cap \{x \in X : f(x) \geq g(x)\}
\]

are \(\mu\)-measurable as well.

**Remark 2.10.** Let \(f : X \to [-\infty, \infty]\). Recall that the positive part of \(f\) is

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

and the negative part is

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

Observe that \(f^+, f^- \geq 0\), \(f = f^+ - f^-\) and \(|f| = f^+ + f^-\). Splitting a function into positive and negative parts will be a useful tool in measure theory. We claim that a function \(f : X \to [-\infty, \infty]\) is \(\mu\)-measurable if and only if \(f^+\) and \(f^-\) are \(\mu\)-measurable.
Reason. \[ \{ x \in X : f^+(x) > a \} = \begin{cases} \{ x \in X : f(x) > a \}, & a > 0, \\ X, & a < 0, \end{cases} \]
is a \( \mu \)-measurable set. Moreover, \( f^- = (-f)^+ \).

**Theorem 2.11.** Assume that \( f, g : X \to [-\infty, \infty] \) are \( \mu \)-measurable functions and \( a \in \mathbb{R} \). Then \( af, f + g, \max(f, g), \min(f, g), \frac{f}{g} \) (\( g \neq 0 \)), are \( \mu \)-measurable functions.

**Warning:** Since functions are extended real-valued, we need to take care about the definitions of \( \max(f, g) \), \( \min(f, g) \), \( f \frac{1}{g} \). The sum is defined outside the bad set

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

but in \( \mathcal{B} \) we have \( \infty - \infty \) situation. We define

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

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

**Proof.** Note that

\[ (f + g)^{-1}((-\infty, a)) = f^{-1}((-\infty, a)) \cup g^{-1}((-\infty, a)) \]

and

\[ (f + g)^{-1}((\infty, a)) = f^{-1}((\infty, a)) \cup g^{-1}((\infty, a)) \]

are \( \mu \)-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 \( \mu \)-measurable. By Lemma 2.8,

\[ \{ x \in X : f(x) + g(x) > a \} = \{ x \in X : g(x) > a - f(x) \} \]
is \( \mu \)-measurable for every \( a \in \mathbb{R} \). This can be also seen directly from

\[ (f + g)^{-1}((-\infty, a)) = \bigcup_{r, s \in \mathbb{R}, r + s < a} (f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, s))). \]

The functions \( |f| \) and \( g^2 \) are \( \mu \)-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

\[ fg = \frac{1}{2} ((f + g)^2 - f^2 - g^2). \]
WARNING: $f^2$ measurable does not imply that $f$ measurable.

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

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

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 $k$th step, we have $2^{k-1}$ open pairwise disjoint intervals $I_{k,i}$, $i = 1, \ldots, 2^{k-1}$. Let $\bar{I}_{m,i}$, $i = 1, \ldots, 2^m - 1$ be the collection of all intervals $I_{k,i}$, with $k \leq m$, $i = 1, \ldots, 2^{k-1}$, organized from left to right. Now we have

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

$k = 0, 1, 2, \ldots$. 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 $x \in \bar{I}_{k,i}$, $i = 1, 2, \ldots, 2^k - 1$ and $f_k$ is linear on $C_k$, $k = 1, 2, \ldots$

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

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

for every $x \in [0, 1]$, the sequence $(f_k)$ 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. Note that $f$ is nondecreasing and is constant on each interval in the complement of the Cantor set. Furthermore, $f : [0, 1] \to [0, 1]$ is onto, that is, $f([0, 1]) = [0, 1]$. The function $f$ 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. The function $f$ is
almost everywhere differentiable and its derivative is zero outside the Cantor set. However, $f$ is not equal to the integral of its derivative, and therefore the fundamental theorem of calculus does not hold.

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

$$C = \bigcap_{k=0}^{\infty} C_k = \bigcap_{k=0}^{\infty} \left[ [0, 1] \setminus \bigcup_{i=1}^{2^k-1} I_{k,i} \right] = [0, 1] \setminus \bigcup_{k=0}^{\infty} \bigcup_{i=1}^{2^k-1} I_{k,i},$$

where $I_{k,i}$ are pairwise disjoint open intervals,

$$m^*(g(C)) = m^* \left( g([0, 1]) \setminus \bigcup_{k=0}^{\infty} \bigcup_{i=1}^{2^k-1} g(I_{k,i}) \right)$$

$$= m^*([0, 2]) - m^* \left( \bigcup_{k=0}^{\infty} \bigcup_{i=1}^{2^k-1} g(I_{k,i}) \right) (g(I_{k,i}) \text{ is an interval})$$

$$= 2 - \lim_{k \to \infty} \sum_{i=1}^{2^k-1} m^*(g(I_{k,i})) (g(I_{k,i}) \text{ are pairwise disjoint})$$

$$= 2 - \lim_{k \to \infty} \sum_{i=1}^{2^k-1} m^*(I_{k,i}) (g(x) = x + a_{k,i} \forall x \in I_{k,i})$$

$$= 2 - m^*([0, 1] \setminus C) = 2 - 1 = 1. \quad (m^*(C) = 0)$$

Figure 2.1: The construction of the Cantor-Lebesgue function.
Thus $g$ maps the zero measure Cantor set $C$ to $g(C)$ set of measure one. Since $m(g(C)) > 0$, by Remark 1.57, there exists $B \subset g(C)$, which is nonmeasurable with respect to the 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 a measurable set $A$ to a nonmeasurable set $B$. Thus measurability is not preserved in continuous mappings.
2. $A$ is a Lebesgue measurable set that is not a Borel set, since a homeomorphism maps Borel sets to Borel sets (exercise).
3. Since $A \subset C$, we conclude that the Cantor set has a subset that is not a Borel set. This shows that the restriction of the Lebesgue measure to the Borel sets is not complete.
4. $\chi_A \circ g^{-1} = \chi_B$, where the function $\chi_B$ is nonmeasurable, but the functions $\chi_A$ and $g^{-1}$ are measurable functions, since $A$ is a measurable set and $g^{-1}$ is continuous.

**Warning:** A composed function of two Lebesgue measurable functions is not necessarily Lebesgue measurable. For positive results, see Remark 2.7.

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

The Cantor-Lebesgue function showed that Lebesgue measurability 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 2.12.** A mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is said to be Lipschitz continuous, if there exists a constant $L$ such that

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

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

**Remark 2.13.** A mapping $f : \mathbb{R}^n \to \mathbb{R}^n$ is of the form $f(x) = (f_1(x), \ldots, f_n(x))$, where $x = (x_1, \ldots, x_n)$ and the coordinate functions $f_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \ldots, n$. Such a mapping $f$ is Lipschitz continuous if and only if all coordinate functions $f_i$, $i = 1, \ldots, n$, satisfy a Lipschitz condition

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

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

**Examples 2.14:**

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

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

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

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

Reason. By the fundamental theorem of calculus,

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

This implies

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

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

Lemma 2.15. Assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a Lipschitz continuous mapping. Then $m^*(f(A)) = 0$ whenever $m^*(A) = 0$.

Theorem: A Lipschitz mapping $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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,

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

for every $y \in B(x_i, r_i)$ and thus $f(y) \in B(f(x_i), Lr_i)$. This implies

$$f(A) \subset f\left(\bigcup_{i=1}^\infty B(x_i, r_i)\right) = \bigcup_{i=1}^\infty f(B(x_i, r_i)) \subset \bigcup_{i=1}^\infty B(f(x_i), Lr_i).$$

By monotonicity, countable subadditivity, translation invariance and the scaling property of the Lebesgue measure we have

$$m^*(f(A)) \leq m^*\left(\bigcup_{i=1}^\infty B(f(x_i), Lr_i)\right) \leq \sum_{i=1}^\infty m^*(B(f(x_i), Lr_i))$$

$$= L^n \sum_{i=1}^\infty m^*(B(x_i, r_i)) < L^n \varepsilon.$$

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

$\square$
Theorem 2.16. Assume that $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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 $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ maps Lebesgue measurable sets to Lebesgue measurable sets.

**Proof.** Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set. By Corollary 1.40 there are compact sets $K_i \subset A$, $i = 1, 2, \ldots$, such that

$$m^* (A \setminus \bigcup_{i=1}^{\infty} K_i) = 0.$$

Since the set $A$ can be written as

$$A = \left( \bigcup_{i=1}^{\infty} K_i \right) \cup \left( A \setminus \bigcup_{i=1}^{\infty} K_i \right),$$

we have

$$f(A) = \bigcup_{i=1}^{\infty} f(K_i) \cup f(A \setminus \bigcup_{i=1}^{\infty} K_i).$$

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, the function $f$ maps sets of measure zero to sets of measure zero. This implies that $f(A \setminus \bigcup_{i=1}^{\infty} K_i)$ is of measure zero and thus
Lebesgue measurable. The set $f(A)$ is measurable as a union of two measurable sets.

**Remark 2.17.** 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

$$m^*(f(A)) \leq cm^*(A)$$

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 1.58.

### 2.5 Limits of measurable functions

Next we show that measurability is preserved under limit operations.

**Theorem 2.18.** Assume that $f_i : X \to [-\infty, \infty]$, $i = 1, 2, \ldots$, are $\mu$-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 $\mu$-measurable functions.

**Remark 2.19.** Recall that

$$\limsup_{i \to \infty} f_i(x) = \inf_{j=1}^{\infty} \sup_{i \geq j} f_i(x)$$

and

$$\liminf_{i \to \infty} f_i(x) = \sup_{j=1}^{\infty} \inf_{i \geq j} f_i(x).$$

**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 $\mu$-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 $\mu$-measurable functions follow immediately.  

\[ \square \]
Theorem 2.20. Assume that \( f_i : X \rightarrow [-\infty, \infty], \ i = 1, 2, \ldots, \) are \( \mu \)-measurable functions. Then

\[
\lim_{i \to \infty} f_i = f
\]

is a \( \mu \)-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

\[
f = \limsup_{i \to \infty} f_i = \liminf_{i \to \infty} f_i
\]

\[\square\]

Example 2.21. Assume that \( f : \mathbb{R} \rightarrow \mathbb{R} \) is differentiable. Then \( f \) is continuous and thus Lebesgue measurable. Moreover, 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 \( f : \mathbb{R} \rightarrow \mathbb{R} \)

\[
f(x) = \begin{cases} 
x^2 \sin \left( \frac{1}{x^2} \right), & x \neq 0, \\
0, & x = 0,
\end{cases}
\]

is differentiable everywhere, but \( f' \) is not continuous at \( x = 0 \).

\[\blacksquare\]

2.6 Almost everywhere

Definition 2.22. Let \( \mu \) be an outer measure on \( X \). A property is said to hold \( \mu \)-almost everywhere \( X \), if it holds in \( X \setminus A \) for a set \( A \subset X \) with \( \mu(A) = 0 \). It is sometimes denoted that the property holds \( \mu \)-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 2.23. Almost everywhere is called “almost surely” in probability theory.

Examples 2.24:
(1) A set $A \subset \mathbb{R}^n$ has Lebesgue measure zero, if for every $\varepsilon > 0$ there exist intervals $I_i, i = 1, 2, \ldots$ such that $A \subset \bigcup_{i=1}^{\infty} I_i$ and
\[
\sum_{i=1}^{\infty} \text{vol}(I_i) < \varepsilon.
\]
The intervals can be replaced by cubes or balls.
(2) Many useful functions as $f : \mathbb{R} \to \mathbb{R}, f(x) = \frac{\sin(x)}{x}$ and $f : \mathbb{R}^n \to \mathbb{R}, f(x) = |x|^{-a}, a > 0,$ are defined only almost everywhere with respect to the Lebesgue measure.

Lemma 2.25. Assume that $f : X \to [-\infty, \infty]$ is a $\mu$-measurable function. If $g : X \to [-\infty, \infty]$ is a function with $f = g$ $\mu$-almost everywhere, then $g$ is a $\mu$-measurable function as well.

**The Moral:** In case $f = g$ $\mu$-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
\[
\{x \in X : g(x) > a\} = ((x \in X : g(x) > a) \cap A) \cup ((x \in X : g(x) > a) \cap (X \setminus A))
\]
\[
= ((x \in X : g(x) > a) \cap A) \cup ((x \in X : f(x) > a) \cap (X \setminus A))
\]
is a $\mu$-measurable set for every $a \in \mathbb{R}$, since $\mu((x \in X : g(x) > a) \cap A) = 0$. □

Remark 2.26. All properties of measurable functions can be relaxed to conditions holding almost everywhere. For example, if $f_i : X \to [-\infty, \infty], i = 1, 2, \ldots,$ are $\mu$-measurable functions and
\[
f = \lim_{i \to \infty} f_i
\]
$\mu$-almost everywhere, then $f$ is a $\mu$-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.
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 2.27.** A function \( f : X \rightarrow \mathbb{R} \) is simple, if its range is a finite set \( \{a_1, \ldots, a_n\} \), \( n \in \mathbb{N} \), and the preimages

\[
 f^{-1}(a_i) = \{ x \in X : f(x) = a_i \}
\]

are \( \mu \)-measurable sets.

**The Moral:** 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 \( A_i = f^{-1}(a_i) \). This implies that a simple function is \( \mu \)-measurable.

**Warning:** 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 \( f : \mathbb{R} \rightarrow \mathbb{R}, f(x) = \chi_Q(x) \) 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.

**Theorem 2.28.** Assume that \( f : X \rightarrow [0, \infty] \) is a nonnegative function. Then \( f \) is a \( \mu \)-measurable function if and only if there exists an increasing sequence \( f_i, i = 1, 2, \ldots \), of simple functions such that

\[
 f(x) = \lim_{i \rightarrow \infty} f_i(x)
\]

for every \( x \in X \).

**The Moral:** Every nonnegative measurable function can be approximated by an increasing sequence of simple functions.

**Proof.** For every \( i = 1, 2, \ldots \) partition \( [0, i) \) into \( i 2^i \) intervals

\[
 I_{i, k} = \left[ \frac{k - 1}{2^i}, \frac{k}{2^i} \right], \quad k = 1, \ldots, i 2^i.
\]
Denote
\[ A_{i,k} = f^{-1}(I_{i,k}) = \{ x \in X : \frac{k-1}{2^i} \leq f(x) < \frac{k}{2^i} \}, \quad k = 1, \ldots, i2^i, \]
and
\[ A_i = f^{-1}([i, \infty)) = \{ x \in X : f(x) \geq i \}. \]
These sets are \( \mu \)-measurable and they form a pairwise disjoint partition of \( X \). The approximating simple function is defined as
\[ f_i(x) = \sum_{k=1}^{i2^i} \frac{k-1}{2^i} \chi_{A_{i,k}}(x) + i \chi_{A_i}(x). \]
Since the sets are pairwise disjoint, \( 0 \leq f_i(x) < f_{i+1}(x) \leq f(x) \) for every \( x \in X \).

In addition,
\[ |f(x) - f_i(x)| \leq \frac{1}{2^i}, \quad \text{if} \quad x \in \bigcup_{k=1}^{i2^i} A_{i,k} = \{ x \in X : f(x) < i \} \]
and
\[ f_i(x) = i, \quad \text{if} \quad x \in A_i = \{ x \in X : f(x) \geq i \}. \]
This implies that
\[ \lim_{i \to \infty} f_i(x) = f(x) \]
for every \( x \in X \).

\( \square \) Follows from the fact that a limit of measurable functions is measurable, see Theorem 2.20.
Remark 2.29. 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.

Now we drop the assumption that the function is nonnegative.

Corollary 2.30. The function \( f : X \rightarrow [-\infty, \infty] \) is a \( \mu \)-measurable function if and
only if there exists a sequence \( f_i, i = 1, 2, \ldots \), of simple functions 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 2.28 there are simple
functions \( g_i \) and \( h_i, i = 1, 2, \ldots \), 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. \( \Box \)
Remarks 2.31:
(1) The sequence \((f_i)\) is increasing, that is, \(|f_i| \leq |f_{i+1}| \leq |f|\) for every \(i = 1, 2, \ldots\), 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 \(f_i : X \to \mathbb{R}, i = 1, 2, \ldots\), 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 \(i_\varepsilon\) such that
\[
|f_i(x) - f(x)| < \varepsilon
\]
whenever \(i \geq i_\varepsilon\). Note that \(i_\varepsilon\) depends on \(x\) and \(\varepsilon\).

(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 \(i_\varepsilon\) such that
\[
|f_i(x) - f(x)| < \varepsilon
\]
for every \(x \in X\) whenever \(i \geq i_\varepsilon\). In this case \(i_\varepsilon\) does not depend on \(x\).

Example 2.32. 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 important in measure theory. We begin with three definitions.

Definition 2.33 (Convergence almost everywhere). Let \(f_i : X \to [-\infty, \infty]\) be a \(\mu\)-measurable function with \(|f_i| < \infty\) \(\mu\)-almost everywhere for every \(i = 1, 2, \ldots\).

We say that \(f_i\) converges to \(f\) almost everywhere in \(X\), if \(f_i(x) \to f(x)\) for \(\mu\)-almost every \(x \in X\).
**CHAPTER 2. MEASURABLE FUNCTIONS**

**Remark 2.34.** Clearly, \( f_i \to f \) almost everywhere in \( X \), if for every \( \varepsilon > 0 \) there is a \( \mu \)-measurable set \( A \subset X \) such that \( \mu(X \setminus A) < \varepsilon \) and \( f_i \to f \) pointwise in \( A \).

**Definition 2.35 (Almost uniform convergence).** Let \( f_i : X \to [-\infty, \infty] \) be a \( \mu \)-measurable function with \( |f_i| < \infty \) \( \mu \)-almost everywhere for every \( i = 1, 2, \ldots \). We say that \( f_i \) converges to \( f \) almost uniformly in \( X \), if for every \( \varepsilon > 0 \) there is a \( \mu \)-measurable set \( A \subset X \) such that \( \mu(X \setminus A) < \varepsilon \) and \( f_i \to f \) uniformly in \( A \).

**The Moral:** Almost uniform convergence is uniform convergence outside a set of arbitrarily small measure.

**Definition 2.36 (Convergence in measure).** Let \( f_i : X \to [-\infty, \infty] \) be a \( \mu \)-measurable function with \( |f_i| < \infty \) \( \mu \)-almost everywhere for every \( i = 1, 2, \ldots \). We say that \( f_i \) converges to \( f \) in measure in \( X \), if

\[
\lim_{i \to \infty} \mu(\{x \in X : |f_i(x) - f(x)| \geq \varepsilon\}) = 0
\]

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 \( i \).

**Remarks 2.37:**

1. The assumption \( |f_i| < \infty \) \( \mu \)-almost everywhere implies that for the limit function we have \( |f| < \infty \) \( \mu \)-almost everywhere.

2. These modes of convergence are not affected by sets of measure zero.

3. Almost uniform convergence implies convergence almost everywhere.

**Reason.** Let \( A_i \) be a measurable set such that \( \mu(A_i) < \frac{1}{i} \) and that \( f_i \to f \) uniformly in \( X \setminus A_i \), \( i = 1, 2, \ldots \). If \( A = \bigcap_{i=1}^{\infty} A_i \), then

\[
\mu(A) \leq \mu(A_i) < \frac{1}{i},
\]

so that \( \mu(A) = 0 \). On the other hand, \( f_i(x) \to f(x) \) for every \( x \in X \setminus A = \bigcup_{i=1}^{\infty} (X \setminus A_i) \). ■

The converse is not true as Example 2.32 shows.
(4) Almost uniform convergence implies convergence in measure.

*Reason.* If \( f_i \to f \) almost uniformly, then for every \( \varepsilon > 0 \) and \( \delta > 0 \), there exists a measurable set \( A \) such that \( \mu(X \setminus A) < \delta \) and \( \sup_{x \in A} |f_i(x) - f(x)| < \varepsilon \), whenever \( i \) is large enough. Thus

\[
\mu(\{x \in X : |f_i(x) - f(x)| \geq \varepsilon\}) \leq \mu(X \setminus A) < \delta,
\]

whenever \( i \) is large enough. This implies that

\[
\lim_{i \to \infty} \mu(\{x \in X : |f_i(x) - f(x)| \geq \varepsilon\}) = 0.
\]

The converse is not true as Example 2.32 shows.

(5) Convergence almost everywhere is called “convergence almost surely” in probability theory and convergence in measure is called “convergence in probability”.

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 2.38:**

1. (Escape to horizontal infinity) Let \( f_i : \mathbb{R} \to \mathbb{R}, \)

\[
f_i(x) = \chi_{[i,i+1]}(x), \quad i = 1, 2, \ldots
\]

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 \( f_i : \mathbb{R} \to \mathbb{R}, \)

\[
f_i(x) = \frac{1}{i} \chi_{[0,i]}(x), \quad i = 1, 2, \ldots
\]

Then \( f_i \to 0 \) uniformly in \( \mathbb{R} \).

3. (Escape to vertical infinity) Let \( f_i : \mathbb{R} \to \mathbb{R}, \)

\[
f_i(x) = i \chi_{[\frac{j}{2^k}, \frac{j+1}{2^k}]}(x), \quad i = 1, 2, \ldots
\]

Then \( f_i \to 0 \) pointwise, almost uniformly and in measure, but not uniformly in \( \mathbb{R} \).

4. (A sliding sequence of functions) Let \( f_j : [0,1] \to \mathbb{R}, \ i = 1, 2, \ldots, \) be defined by

\[
f_{2^k+j}(x) = k \chi_{\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]}(x), \quad k = 0, 1, 2, \ldots, \ j = 0, 1, \ldots, 2^k - 1.
\]

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,

\[
m^*([x \in [0,1] : f_{2^k+j}(x) \geq \varepsilon]) = m^*\left(\left[\frac{j}{2^k}, \frac{j+1}{2^k}\right]\right) = \frac{1}{2^k} \to 0 \quad \text{as} \quad k \to \infty.
\]
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.

**Figure 2.5:** A sliding sequence of functions.

The next result shows that a converging subsequence, as in the sliding sequence of functions above, always exists.

**Theorem 2.39.** Assume that \( f_i \to f \) in measure. Then there exists a subsequence \((f_{i_k})\) such that \( f_{i_k} \to f \mu\)-almost everywhere.

**Remark 2.40.** This implies that the limit function is measurable and unique.

**Proof.** Choose \( i_1 \) such that

\[
\mu(\{x \in X : |f_{i_1}(x) - f(x)| \geq 1\}) < \frac{1}{2}.
\]

Assume then that \( i_1, \ldots, i_k \) have been chosen. Choose \( i_{k+1} > i_k \) such that

\[
\mu\left(\left\{x \in X : |f_{i_{k+1}}(x) - f(x)| \geq \frac{1}{k+1}\right\}\right) < \frac{1}{2^{k+1}}.
\]

Define

\[
A_j = \bigcup_{k=j}^{\infty} \left\{x \in X : |f_{i_k}(x) - f(x)| \geq \frac{1}{k+1}\right\}, \quad j = 1, 2, \ldots,
\]

Clearly \( A_{j+1} \subset A_j \) and denote \( A = \bigcap_{j=1}^{\infty} A_j \). Then for every \( j = 1, 2, \ldots, \) we have

\[
\mu(A) \leq \mu(A_j) \leq \sum_{k=j}^{\infty} \frac{1}{2^k} = \frac{2}{2^j}.
\]
and thus $\mu(A) = 0$. Since

$$X \setminus A = \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} \left\{ x \in X : |f_{i_k}(x) - f(x)| < \frac{1}{k} \right\},$$

for every $x \in X \setminus A$ there exists $j$ such that

$$|f_{i_k}(x) - f(x)| < \frac{1}{k} \quad \text{for every} \quad k \geq j.$$

This implies that

$$|f_{i_k}(x) - f(x)| \to 0$$

as $k \to \infty$ for every $x \in X \setminus A$. \qed

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{comparison_modes_convergence.png}
\caption{Comparison of modes of convergence.}
\end{figure}

\section{Egoroff’s and Lusin’s theorems}

The next result gives the main motivation for almost uniform convergence.

\textbf{Theorem 2.41 (Egoroff's theorem).} Assume that $\mu(X) < \infty$. Let $f_i : X \to [-\infty, \infty]$ be a $\mu$-measurable function with $|f_i| < \infty$ $\mu$-almost everywhere for every $i = 1, 2, \ldots$ such that $f_i \to f$ almost everywhere in $X$ and $|f| < \infty$ $\mu$-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 2.42.** The sequence \( f_i = \chi_{B(0,i)} \), \( i = 1, 2, \ldots \), converges pointwise to \( f = 1 \) in \( \mathbb{R}^n \), but it does not converge uniformly outside any bounded set. On the other hand, if \( |f_x| < \infty \) 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

\[
A_{j, k} = \bigcup_{i=j}^{\infty} \left\{ x \in X : |f_i(x) - f(x)| > \frac{1}{2^k} \right\}, \quad j, k = 1, 2, \ldots
\]

Then \( A_{j+1, k} \subset A_{j, k} \) for every \( j, k = 1, 2, \ldots \). 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

\[
\mu(A_{j, k}) < \frac{\varepsilon}{2^k + 1}.
\]

Denote \( A = X \setminus \bigcup_{k=1}^{\infty} A_{j, k} \). This implies

\[
\mu(X \setminus A) \leq \sum_{k=1}^{\infty} \mu(A_{j, k}) \leq \sum_{k=1}^{\infty} \frac{\varepsilon}{2^k + 1} < \varepsilon.
\]

Then for every \( k = 1, 2, \ldots \) and \( i \geq j_k \) we have

\[
|f_i(x) - f(x)| \leq \frac{1}{2^k}
\]

for every \( x \in A \). This implies that \( f_i \to f \) uniformly in \( A \). \( \square \)

**Remark 2.43.** If \( \mu(X) = \infty \), we can apply Egoroff's theorem for \( \mu \)-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 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 \( i = 1, 2, \ldots \) 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 \( \varepsilon \) and \( f_i \to f \) locally uniformly in \( \mathbb{R}^n \setminus A \).

**Remark 2.44.** Relations of measurable sets and functions to standard open sets and continuous functions are summarized in Littlewood's three principles.
2. MEASURABLE FUNCTIONS 72

(1) Every measurable set is almost open (Theorem 1.37).
(2) Pointwise convergence is almost uniform (Egoroff’s theorem 2.41).
(3) A measurable function is almost continuous (Lusin’s theorem 2.45).

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 2.45 (Lusin’s theorem).** Let \( \mu \) be a Borel regular outer measure on \( \mathbb{R}^n \), \( A \subset \mathbb{R}^n \) a \( \mu \)-measurable set such that \( \mu(A) < \infty \) and \( f : \mathbb{R}^n \to [-\infty, \infty] \) be a \( \mu \)-measurable function such that \( |f| < \infty \) \( \mu \)-almost everywhere. Then 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 2.46:**

(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

\[
0 \leq \mu(N) = \mu(A \setminus B) = \mu\left( A \setminus \bigcup_{i=1}^{\infty} K_i \right)
= \mu\left( \bigcup_{i=1}^{\infty} (A \setminus K_i) \right) \leq \mu(A \setminus K_i) < \frac{1}{i}
\]

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 \( \mu \)-measurable, since \( f \) is continuous in \( B \) and \( \{ x \in A \setminus B : f(x) > a \} \) is \( \mu \)-measurable, since it is a set of measure zero. This implies that \( f \) is \( \mu \)-measurable in \( A \).

**Proof.** For every \( i = 1, 2, \ldots \), let \( B_{i,j}, \ j = 1, 2, \ldots \), be disjoint Borel sets such that

\[
\bigcup_{j=1}^{\infty} B_{i,j} = \mathbb{R} \quad \text{and} \quad \text{diam}(B_{i,j}) < \frac{1}{i}.
\]
Denote $A_{i,j} = A \cap f^{-1}(B_{i,j})$. By Lemma 2.5, the set $A_{i,j}$ is $\mu$-measurable. Moreover

\[ A = A \cap f^{-1}(\mathbb{R}) = A \cap f^{-1} \left( \bigcup_{j=1}^{\infty} B_{i,j} \right) = \bigcup_{j=1}^{\infty} A_{i,j}, \quad i = 1, 2, \ldots \]

Since $\mu(A) < \infty$, $\nu = \mu|A$ is a Radon measure by Lemma 1.35. By Corollary 1.40, 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

\[ \mu \left( A \setminus \bigcup_{j=1}^{\infty} K_{i,j} \right) = \nu \left( A \setminus \bigcup_{j=1}^{\infty} K_{i,j} \right) = \nu \left( \bigcup_{j=1}^{\infty} A_{i,j} \setminus \bigcup_{j=1}^{\infty} K_{i,j} \right) \leq \nu \left( \bigcup_{j=1}^{\infty} (A_{i,j} \setminus K_{i,j}) \right) \leq \sum_{j=1}^{\infty} \nu(A_{i,j} \setminus K_{i,j}) < \frac{\varepsilon}{2^i}. \]

Since $\mu(A) < \infty$, Theorem 1.13 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

\[ g_i(x) = \alpha_{i,j}, \quad \text{when} \quad x \in K_{i,j}, \quad j = 1, \ldots, k_i. \]

Since $K_{i,1}, \ldots, K_{i,k_i}$ are pairwise disjoint compact sets,

\[ \text{dist}(K_{i,j}, K_{i,l}) > 0 \quad \text{when} \quad j \neq l. \]

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 $\text{diam}(B_{i,j}) < \frac{1}{i}$. The set $K = \bigcup_{i=1}^{\infty} K_i$ is compact and

\[ \mu(A \setminus K) = \mu \left( A \setminus \bigcup_{i=1}^{\infty} K_i \right) = \mu \left( \bigcup_{i=1}^{\infty} (A \setminus K_i) \right) \leq \sum_{i=1}^{\infty} \mu(K_i) < \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} = \varepsilon. \]

Since

\[ |f(x) - g_i(x)| < \frac{1}{i} \quad \text{for every} \quad x \in K, \quad i = 1, 2, \ldots, \]

we see that $g_i \to f$ uniformly in $K$. The function $f$ is continuous in $K$ as a uniform limit of continuous functions. □
WARING: Note carefully, that $f|_K$ denotes the restriction of $f$ to $K$. Theorem 2.45 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. $f : [0, 1] \to \mathbb{R}, f(x) = \chi_{[0, 1]}(x)$ is discontinuous at every point of $[0, 1]$. However, $f|_{[0, 1] \cap Q} = 1$ and $f|_{[0, 1] \setminus Q} = 0$ 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 2.47.** Let $\mu$ be a Borel regular outer measure on $\mathbb{R}^n$, $A \subset \mathbb{R}^n$ a $\mu$-measurable set such that $\mu(A) < \infty$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a $\mu$-measurable function such that $|f| < \infty$ $\mu$-almost everywhere. Then for every $\varepsilon > 0$ there exists a continuous function $\overline{f} : \mathbb{R}^n \to \mathbb{R}$ such that

$$\mu(\{x \in A : \overline{f}(x) \neq f(x)\}) < \varepsilon.$$ 

WARING: The corollary does not imply that there is a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f(x) = f(x)$ $\mu$-almost everywhere, see Example 3.33.

**Proof:** Let $\varepsilon > 0$. By Lusin’s theorem 2.45, there exists a compact set $K \subset A$ such that $\mu(A \setminus K) < \varepsilon$ and $f|_K$ is continuous. Then 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 do not prove Tieze’s theorem in this course. Then

$$\mu(\{x \in A : \overline{f}(x) \neq f(x)\}) < \mu(A \setminus K) < \varepsilon,$$

which implies the claim. \hfill \Box

**Remarks 2.48:**

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 $|f(x)| \leq M$ for every $x \in F$, then if $|\overline{f}(x)| \leq M$ for every $x \in X$.

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 2.49.** Let $A \subset \mathbb{R}^n$ be a Lebesgue measurable set with $m(A) < \infty$ and $f = \chi_A$. By Theorem 1.52, 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 1.22, define

$$\overline{f}(x) = \frac{\text{dist}(x, \mathbb{R}^n \setminus G)}{\text{dist}(x, \mathbb{R}^n \setminus G) + \text{dist}(x, K)}.$$
Then $\overline{f}$ is a continuous function in $\mathbb{R}^n$ and
\[
m((x \in \mathbb{R}^n : \overline{f}(x) \neq f(x))) \leq m(G \setminus K) = m(G \setminus A) + m(A \setminus K) < \varepsilon.
\]

In this special case, the function in the previous corollary can be constructed explicitly.
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 $\mu$-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 $\{a_1, \ldots, a_n\}, n \in \mathbb{N}$, and the preimages $f^{-1}(\{a_i\}) = \{x \in X : f(x) = a_i\}$ are $\mu$-measurable sets, see Definition 2.27. 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

$$f = \sum_{i=1}^n a_i \chi_{A_i}, \quad n \in \mathbb{N},$$

where $A_i = f^{-1}(\{a_i\})$. This implies that a simple function is $\mu$-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 $\mu$ 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 \( f : \mathbb{R} \to \mathbb{R}, f(x) = \chi_Q(x) \) is simple with respect to the one dimensional Lebesgue measure and \( \int f(x) \, dx = 0 \).

**Remarks 3.3:**
1. \( 0 \leq \int_X f \, d\mu \leq \infty \).
2. If \( f \) is a simple function and \( A \) is \( \mu \)-measurable subset of \( X \), then \( f \chi_A \) is a simple function.
3. (Compatibility with the measure) If \( A \) is \( \mu \)-measurable subset of \( X \), then \( \int_X \chi_A \, d\mu = \mu(A) \).
4. If \( A \) is a \( \mu \)-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. We leave this as an exercise.
5. If \( f \) and \( g \) are nonnegative simple functions such that \( f = g \mu \)-almost everywhere, then \( \int_X f \, d\mu = \int_X g \, d\mu \).
6. \( \int_X f \, d\mu = 0 \) if and only if \( f = 0 \mu \)-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 \( \mu \)-measurable sets with \( A \subset B \), then \( \int_A f \, d\mu \leq \int_B f \, d\mu \).
2. (Homogeneity) \( \int_X af \, d\mu = a \int_X f \, d\mu, a \geq 0 \).
3. (Linearity) \( \int_X (f + g) \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu \).
4. (Monotonicity in functions) \( f \leq g \) implies \( \int_X f \, d\mu \leq \int_X g \, d\mu \).

**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
   \[
   C_{i,j} = A_i \cap B_j, \quad i = 1, \ldots, n, j = 1, \ldots, m,
   \]
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
\[
\int_{X} (f + g) \, d\mu = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i + b_j) \mu(C_{i,j})
\]
\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(A_i \cap B_j) + \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(A_i \cap B_j)
\]
\[
= \sum_{i=1}^{n} a_i \mu(A_i) + \sum_{j=1}^{m} b_j \mu(B_j)
\]
\[
= \int_{X} f \, d\mu + \int_{X} g \, d\mu.
\]
To prove (4) we note that on the sets \( C_{i,j} = A_i \cap B_j \) we have \( f = a_i \leq b_j = g \) and thus
\[
\int_{X} f \, d\mu = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i \mu(C_{i,j}) \leq \sum_{i=1}^{n} \sum_{j=1}^{m} b_j \mu(C_{i,j}) = \int_{X} g \, d\mu.
\]

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 \( A_i, i = 1, 2, \ldots \) are \( \mu \)-measurable sets, then
\[
\int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu \leq \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu
\]
and
\[
\int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu
\]
if the sets \( A_i, i = 1, 2, \ldots \), are pairwise disjoint. Moreover, if \( A_i \supset A_{i+1} \) for every \( i \) and \( \int_{A_i} f \, d\mu < \infty \), 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 \( \mu \)-measurable function. The integral of \( f \) with respect to \( \mu \) is
\[
\int_X f \, d\mu = \sup \left\{ \int_X g \, d\mu : g \text{ is simple and } 0 \leq g(x) \leq f(x) \text{ for every } x \in X \right\}.
\]
A nonnegative function is called integrable, if
\[
\int_X f \, d\mu < \infty.
\]

**The Moral:** Integral is defined for all nonnegative measurable functions. Observe, that the integral may be \( \infty \).

**Remarks 3.7:**
1. As before, if \( A \) is a \( \mu \)-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 \( \mu \)-measurable functions.

1. (Monotonicity in sets) If \( A \) and \( B \) are \( \mu \)-measurable sets with \( A \subset B \), then
\[
\int_A f \, d\mu \leq \int_B f \, d\mu.
\]
2. (Homogeneity) \( \int_X af \, d\mu = a \int_X f \, d\mu, \ a \geq 0 \).
3. (Linearity) \( \int_X (f + g) \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu \).
4. (Monotonicity in functions) \( f \leq g \) implies \( \int_X f \, d\mu \leq \int_X g \, d\mu \).
5. (Tchebyshev’s inequality)
\[
\mu(\{x \in X : f(x) > a\}) \leq \frac{1}{a} \int_X f \, d\mu
\]
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. (1) Follows immediately from the corresponding property for nonnegative simple functions.

(2) If \( a = 0 \), then
\[
\int_X (0 f) \, d\mu = \int_X 0 \, d\mu = 0 = \int_X f \, d\mu.
\]
Let then \( a > 0 \). If \( g \) is simple and \( 0 \leq g \leq f \), then \( a g \) is a nonnegative simple function with \( a g \leq af \). It follows that
\[
a \int_X g \, d\mu = \int_X a g \, d\mu \leq \int_X af \, d\mu.
\]
Taking the supremum over all such functions \( g \) implies
\[
a \int_X f \, d\mu \leq \int_X af \, d\mu.
\]
Applying this inequality gives
\[
a \int_X af \, d\mu = a \left( \frac{1}{a} \int_X af \, d\mu \right) \leq a \int_X (af) \, d\mu = a \int_X f \, d\mu.
\]
(3) Exercise, see also the remark after the monotone convergence theorem.

(4) Let \( h \) be a simple function with \( 0 \leq h(x) \leq f(x) \) for every \( x \in X \). Then \( 0 \leq h(x) \leq g(x) \) for every \( x \in X \) and thus \( \int_X h \, d\mu \leq \int_X g \, d\mu \). By taking supremum over all such functions \( h \) we have \( \int_X f \, d\mu \leq \int_X g \, d\mu \).

(5) Since \( f \geq a \chi_{\{x \in X : f(x) > a\}} \), we have
\[
a \mu(\{x \in X : f(x) > a\}) = \int_X a \chi_{\{x \in X : f(x) > a\}} \, d\mu \leq \int_X f \, d\mu.
\]
\( \square \)

Lemma 3.9. Let \( f : X \to [0, \infty] \) be a \( \mu \)-measurable function.

(1) (Vanishing) \( \int_X f \, d\mu = 0 \) if and only if \( f = 0 \) \( \mu \)-almost everywhere.

(2) (Finiteness) \( \int_X f \, d\mu < \infty \) implies \( f < \infty \) \( \mu \)-almost everywhere.

Warning: The claim (1) is not necessarily true for a sign changing function. The converse of claim (2) is not true: \( f < \infty \) \( \mu \)-almost everywhere does not imply that \( \int_X f \, d\mu < \infty \).

Proof. (1)\( \Rightarrow \) Let
\[
A_i = \left\{ x \in X : f(x) > \frac{1}{i} \right\}, \quad i = 1, 2, \ldots
\]
By Tchebyshev's inequality
\[
0 \leq \mu(A_i) \leq i \int_X f \, d\mu = 0
\]
which implies that \( \mu(A_i) = 0 \) for every \( i = 1, 2, \ldots \). Thus
\[
\mu(\{ x \in X : f(x) > 0 \}) = \mu \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu(A_i) = 0.
\]
\( \Rightarrow \) Since \( \mu(\{ x \in X : f(x) > 0 \}) = 0 \), we have
\[
0 = \int_{\{ x \in X : f(x) > 0 \}} \infty \, d\mu = \int_X \infty \chi_{\{ x \in X : f(x) > 0 \}} \, d\mu \geq \int_X f \, d\mu \geq 0.
\]

Thus \( \int_A f \, d\mu = 0 \). Another way to prove this claim is to use the definition of integral directly (exercise).

By Tchebychev’s inequality
\[
\mu(\{ x \in X : f(x) = \infty \}) \leq \mu(\{ x \in X : f(x) > i \}) \leq \frac{1}{i} \int_X f \, d\mu \to 0
\]
as \( i \to \infty \) because \( \int_X f \, d\mu < \infty \).

**Lemma 3.10.** Let \( f, g : X \to [0, \infty] \) be \( \mu \)-measurable functions. If \( f = g \) \( \mu \)-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
\[
\int_N f \, d\mu = \int_N g \, d\mu.
\]

It follows that
\[
\int_X f \, d\mu = \int_{X \setminus N} f \, d\mu + \int_N f \, d\mu = \int_{X \setminus N} f \, d\mu = \int_{X \setminus N} g \, d\mu + \int_N g \, d\mu = \int_X g \, d\mu. \quad \Box
\]

### 3.3 Monotone convergence theorem

Assume that \( f_i : X \to [0, \infty], \ i = 1, 2, \ldots, \) are \( \mu \)-measurable functions and that \( f_i \to f \) either everywhere or \( \mu \)-almost everywhere as \( i \to \infty \). Then \( f \) is a \( \mu \)-measurable function. 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 $f_i : \mathbb{R} \rightarrow \mathbb{R}$,
   
   $$f_i(x) = \chi_{[i,i+1]}(x), \quad i = 1, 2, \ldots$$
   
   Then $f_i(x) = 0$ as $i \rightarrow \infty$ for every $x \in \mathbb{R}$, but
   
   $$\int_{\mathbb{R}} \lim_{i \rightarrow \infty} f_i \, dm = \int_{\mathbb{R}} 0 \, dm = 0 < 1 = \lim_{i \rightarrow \infty} \int_{\mathbb{R}} f_i \, dm.$$  

2. (Escape to width infinity) Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$,
   
   $$f_i(x) = \frac{1}{i} \chi_{[0,i]}(x), \quad i = 1, 2, \ldots$$
   
   Then $f_i \rightarrow 0$ uniformly in $\mathbb{R}$, but
   
   $$\int_{\mathbb{R}} \lim_{i \rightarrow \infty} f_i \, dm = \int_{[0,\infty)} 0 \, dm = 0 < 1 = \lim_{i \rightarrow \infty} \int_{\mathbb{R}} f_i \, dm.$$  

3. (Escape to vertical infinity) Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$,
   
   $$f_i(x) = i \chi_{[1/i,2/i]}(x), \quad i = 1, 2, \ldots$$
   
   Then $\lim_{i \rightarrow \infty} f_i(x) = 0$ for every $x \in \mathbb{R}$, but
   
   $$\int_{\mathbb{R}} \lim_{i \rightarrow \infty} f_i \, dm = \int_{\mathbb{R}} 0 \, dm = 0 < 1 = \lim_{i \rightarrow \infty} \int_{\mathbb{R}} f_i \, dm.$$  

Observe that the sequence is not increasing.

4. Let $f_i : \mathbb{R} \rightarrow \mathbb{R}$, $f_i(x) = \frac{1}{i}$, $i = 1, 2, \ldots$ Then $\lim_{i \rightarrow \infty} f_i(x) = 0$ for every $x \in \mathbb{R}$, but
   
   $$\int_{\mathbb{R}} \lim_{i \rightarrow \infty} f_i \, dm = 0 < \infty = \lim_{i \rightarrow \infty} \int_{\mathbb{R}} f_i \, dm.$$  

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 $f_i : X \rightarrow [0,\infty]$ are $\mu$-measurable functions such that $f_i \leq f_{i+1}$, $i = 1, 2, \ldots$, then

$$\int_X \lim_{i \rightarrow \infty} f_i \, d\mu = \lim_{i \rightarrow \infty} \int_X f_i \, d\mu.$$  

The moral: The order of taking limit and integral can be switched for an increasing sequence of nonnegative measurable functions.

Remarks 3.13:

1. The limit function $f = \lim_{i \rightarrow \infty} f_i$ is measurable as a pointwise limit of measurable functions.

2. It is enough to assume that $f_i \leq f_{i+1}$ almost everywhere.
(3) The limits may be infinite.

(4) In the special case when \( f_i = \chi_{A_i} \), where \( A_i \) is \( \mu \)-measurable and \( A_i \subset A_{i+1} \), the monotone convergence theorem reduces to the upwards monotone convergence result for measures, see Theorem 1.13.

**Proof.** Let \( f = \lim_{i \to -\infty} f_i \). By monotonicity

\[ \int_X f_i \, d\mu \leq \int_X f_{i+1} \, d\mu \leq \int_X f \, d\mu \]

for every \( i = 1, 2, \ldots \) This implies that the limit exists and

\[ \lim_{i \to -\infty} \int_X f_i \, d\mu \leq \int_X f \, d\mu. \]

To prove the reverse inequality, let \( g \) be a nonnegative simple function with \( g \leq f \). Let \( 0 < t < 1 \) and

\[ A_i = \{ x \in X : f_i(x) \geq tg(x) \}, \quad i = 1, 2, \ldots \]

By Lemma 2.8 and Remark 2.8, the sets \( A_i \) are \( \mu \)-measurable and \( A_i \subset A_{i+1} \), \( i = 1, 2, \ldots \).

**Claim:** \( \bigcup_{i=1}^{\infty} A_i = X \).

**Reason.** Since \( A_i \subset X \), \( i = 1, 2, \ldots \), we have \( \bigcup_{i=1}^{\infty} A_i \subset X \).

For every \( x \in X \), either \( f(x) \leq tg(x) \) or \( f(x) > tg(x) \). If \( f(x) \leq tg(x) \), then \( f(x) \leq tg(x) \leq tf(x) \) and, since \( 0 < t < 1 \), we have \( f(x) = 0 \). In this case \( x \in A_i \) for every \( i = 1, 2, \ldots \). 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

\[ \int_X f_i \, d\mu \leq \int_{A_i} f_i \, d\mu \leq \int_{A_i} tg \, d\mu = t \int_{A_i} g \, d\mu \rightarrow t \int_X g \, d\mu \]

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

\[ \lim_{i \to -\infty} \int_X f_i \, d\mu \geq t \int_X f \, d\mu. \]

By taking the supremum over all nonnegative simple functions \( g \) we have

\[ \lim_{i \to -\infty} \int_X f_i \, d\mu \geq t \int_X f \, d\mu \]

and the claim follows by passing \( t \to 1 \). □

**Remarks 3.14:**

(1) By Theorem 2.28 for every nonnegative \( \mu \)-measurable function \( f \) there is an increasing sequence \( f_i, i = 1, 2, \ldots \), 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 $f_i, i = 1, 2, \ldots,$ are nonnegative simple functions such that $f_i \leq f_{i+1}$ 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 $\mu$-measurable functions. Let $f_i, i = 1, 2, \ldots$, be an increasing sequence of nonnegative simple functions such that $f = \lim_{i \to \infty} f_i$ and let $g_i, i = 1, 2, \ldots$, 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

$$\int_X (f + g) \, d\mu = \lim_{i \to \infty} \int_X (f_i + g_i) \, d\mu = \lim_{i \to \infty} \int_X f_i \, d\mu + \lim_{i \to \infty} \int_X g_i \, d\mu = \int_X f \, d\mu + \int_X g \, d\mu.$$ 

This shows the approximation by simple functions can be used to prove properties of the integral, compare to Lemma 3.8.

**Corollary 3.15.** Let $f_i : X \to [0, \infty], i = 1, 2, \ldots,$ be nonnegative $\mu$-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 $s_n = f_1 + \cdots + f_n = \sum_{i=1}^{n} f_i$ be the $n$th partial sum and

$$f = \lim_{n \to \infty} s_n = \sum_{i=1}^{\infty} f_i.$$ 

The functions $s_n, n = 1, 2, \ldots,$ form an increasing sequence of nonnegative $\mu$-measurable functions. By the monotone convergence theorem

$$\int_X f \, d\mu = \lim_{n \to \infty} \int_X s_n \, d\mu = \lim_{n \to \infty} \sum_{i=1}^{n} \int_X f_i \, d\mu = \sum_{i=1}^{\infty} \int_X f_i \, d\mu.$$
Remark 3.16. Let \( f \) be a nonnegative \( \mu \)-measurable function on \( X \). Define

\[
\nu(A) = \int_A f \, d\mu
\]

for any \( \mu \)-measurable set \( A \). Then \( \nu \) is a measure.

Reason. It is clear that \( \nu \) is nonnegative and that \( \nu(\emptyset) = 0 \). We show that \( \nu \) is countably additive on pairwise disjoint \( \mu \)-measurable sets. Let \( A_i, i = 1, 2, \ldots \), be pairwise disjoint \( \mu \)-measurable sets and let \( f_i = f \chi_{A_i} \). Then by the previous corollary,

\[
\sum_{i=1}^{\infty} \nu(A_i) = \sum_{i=1}^{\infty} \int_X f_i \, d\mu = \int_X \sum_{i=1}^{\infty} f_i \, d\mu = \int_X \sum_{i=1}^{\infty} f \chi_{A_i} \, d\mu = \int_X f \sum_{i=1}^{\infty} \chi_{A_i} \, d\mu = \int_X f \chi_{\bigcup_{i=1}^{\infty} A_i} \, d\mu = \int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu = \nu\left(\bigcup_{i=1}^{\infty} A_i\right).
\]

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 3.28.

1. (Countable additivity) If \( A_i, i = 1, 2, \ldots \) are pairwise disjoint \( \mu \)-measurable sets, then

\[
\int_{\bigcup_{i=1}^{n} A_i} f \, d\mu = \sum_{i=1}^{n} \int_{A_i} f \, d\mu.
\]

2. (Countable subadditivity) If \( A_i, i = 1, 2, \ldots \) are \( \mu \)-measurable sets, then

\[
\int_{\bigcup_{i=1}^{n} A_i} f \, d\mu \leq \sum_{i=1}^{n} \int_{A_i} f \, d\mu.
\]

3. (Downwards monotone convergence) If \( A_i, i = 1, 2, \ldots \), are \( \mu \)-measurable, \( A_i \supset A_{i+1} \) for every \( i \) and \( \int_{A_i} f \, d\mu < \infty \), then

\[
\int_{\bigcap_{i=1}^{n} A_i} f \, d\mu = \lim_{i \to \infty} \int_{A_i} f \, d\mu.
\]

4. (Upwards monotone convergence) If \( A_i, i = 1, 2, \ldots \), are \( \mu \)-measurable and \( A_i \supset A_{i+1} \) for every \( i \), then

\[
\int_{\bigcup_{i=1}^{n} A_i} f \, d\mu = \lim_{i \to \infty} \int_{A_i} f \, d\mu.
\]

### 3.4 Fatou’s lemma

The next convergence result holds without monotonicity assumptions.
Theorem 3.17 (Fatou’s lemma). If $f_i : X \to [0, \infty]$, $i = 1, 2, \ldots$, are $\mu$-measurable functions, then

$$\int_X \liminf_{i \to \infty} f_i \, d\mu \leq \liminf_{i \to \infty} \int_X f_i \, d\mu.$$ 

**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 3.18:

1. The power of Fatou’s lemma is that there are no assumptions on the convergences. In particular, the limits

$$\lim_{i \to \infty} f_i \text{ and } \lim_{i \to \infty} \int_X f_i \, d\mu$$

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

$$\liminf_{i \to \infty} f_i(x) = \sup_{i \to \infty} \inf_{j \to \infty} f_j(x) = \lim_{j \to \infty} \inf_{i \to \infty} f_i(x),$$

where $g_j = \inf_{i \to \infty} f_i$. The functions $g_j$, $j = 1, 2, \ldots$, form an increasing sequence of $\mu$-measurable functions. By the monotone convergence theorem

$$\int_X \liminf_{i \to \infty} f_i \, d\mu = \int_X \lim_{j \to \infty} g_j \, d\mu$$

$$= \lim_{j \to \infty} \int_X g_j \, d\mu$$

$$\leq \liminf_{i \to \infty} \int_X f_i \, d\mu,$$

where the last inequality follows from the fact that $g_i \leq f_i$. 

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 $f^+, f^- \geq 0$ and $f = f^+ - f^-$. 

**Definition 3.19.** Let $f : X \to [-\infty, \infty]$ be a $\mu$-measurable function. If either $\int_X f^- \, d\mu < \infty$ or $\int_X f^+ \, d\mu < \infty$, then the integral if $f$ 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 $f \in L^1(X; \mu)$. 

CHAPTER 3. INTEGRATION

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 3.20. (Triangle inequality) A function $f$ is integrable if and only if $|f|$ is integrable, that is, $\int_X |f| d\mu < \infty$. In this case,

$$\left| \int_X f d\mu \right| \leq \int_X |f| d\mu.$$

Reason. 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 $X$.

Assume that $|f|$ is integrable in $X$. Since $0 \leq f^+ \leq |f|$ and $0 \leq f^- \leq |f|$ we have

$$\int_X f^+ d\mu \leq \int_X |f| d\mu < \infty \quad \text{and} \quad \int_X f^- d\mu \leq \int_X |f| d\mu < \infty.$$

It follows that $f$ is integrable in $X$.

Moreover,

$$\left| \int_X f d\mu \right| = \left| \int_X f^+ d\mu - \int_X f^- d\mu \right| = \left| \int_X f^+ d\mu + \int_X f^- d\mu \right|$$

$$= \int_X f^+ d\mu + \int_X f^- d\mu = \int_X (f^+ + f^-) d\mu = \int_X |f| d\mu. \quad \blacksquare$$

Remarks 3.21:

(1) If $f \in L^1(X;\mu)$, then $|f| \in L^1(X;\mu)$ and by Lemma 3.9 we have $|f| < \infty$ $\mu$-almost everywhere in $X$.

(2) (Majorant principle) Let $f : X \to [-\infty, \infty]$ be a $\mu$-measurable function. If there exists a nonnegative integrable function $g$ such that $|f| \leq g$ $\mu$-almost everywhere, then $f$ is integrable.

Reason. $\int_X |f| d\mu \leq \int_X g d\mu < \infty. \quad \blacksquare$

(3) A measure space $(X, \mathcal{M}, \mu)$ with $\mu(X) = 1$ is called a probability or sample space, $\mu$ 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 3.22:

1. Let $X = \mathbb{R}^n$ and $\mu$ be the Lebesgue measure. We shall discuss properties of the Lebesgue measure in detail later.

2. Let $X = \mathbb{N}$ and $\mu$ be the counting measure. Then all functions are $\mu$-measurable. Observe that a function $f : \mathbb{N} \to \mathbb{R}$ is a sequence of real numbers with $x_i = f(i)$, $i = 1, 2, \ldots$. 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 $\mu$-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 3.23. 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) $f \leq g$ implies $\int_X f \, d\mu \leq \int_X g \, d\mu$.
4. (Vanishing) $\mu(X) = 0$ implies $\int_X f \, d\mu = 0$.
5. (Almost everywhere equivalence) If $f = g \, \mu$-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 3.24. Since $f, g \in L^1(X; \mu)$, we have $|f| < \infty$ and $|g| < \infty$ $\mu$-almost everywhere in $X$. Thus $f + g$ is defined $\mu$-almost everywhere in $X$. 


CHAPTER 3. INTEGRATION

Proof. (1) If \( a > 0 \), 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.

(2) Let \( h = f + g \). Then \( h \) is defined almost everywhere and measurable. The pointwise inequality

\[
|h| \leq |f| + |g|
\]

implies

\[
\int_X |h| \, d\mu \leq \int_X |f| \, d\mu + \int_X |g| \, d\mu < \infty
\]

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

\[
\int_X h \, d\mu = \int_X h^+ \, d\mu - \int_X h^- \, d\mu
\]

\[
= \int_X f^+ \, d\mu - \int_X f^- \, d\mu + \int_X g^+ \, d\mu - \int_X g^- \, d\mu
\]

\[
= \int_X f \, d\mu + \int_X g \, d\mu.
\]

(3) (1) and (2) imply that \( g - f \geq 0 \) is integrable and

\[
\int_X g \, d\mu = \int_X f \, d\mu + \int_X (g - f) \, d\mu \geq \int_X f \, d\mu.
\]

(4) \( \mu(X) = 0 \) implies \( \int_X f^+ \, d\mu = 0 \) and \( \int_X f^- \, d\mu = 0 \) and consequently \( \int_X f \, d\mu = 0 \).

(5) If \( f = g \) \( \mu \)-almost everywhere in \( X \), then \( f^+ = g^+ \) and \( f^- = g^- \) \( \mu \)-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. \( \square \)
3.6 Dominated convergence theorem

Now we are ready to state the principal convergence theorem in the theory of integration.

Theorem 3.25 (Dominated convergence theorem). Let \( f_i : X \to [-\infty, \infty], i = 1, 2, \ldots \), be \( \mu \)-measurable functions such that \( f_i \to f \) \( \mu \)-almost everywhere as \( i \to \infty \).

If there exists an integrable function \( g \) such that \( |f_i| \leq g \) \( \mu \)-almost everywhere for every \( i = 1, 2, \ldots \), 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 3.26. As the moving bump examples show, see Examples 3.11, assumption about the integrable majorant is necessary.

Proof. Let

\[
N = \{ x \in X : \liminf_{i \to \infty} f_i(x) \neq f(x) \} \cup \{ x \in X : \limsup_{i \to \infty} f_i(x) \neq f(x) \}
\]

\[ \cup \bigcup_{i=1}^{\infty} \{ x \in X : |f_i(x)| > g(x) \}. \]

Then \( \mu(N) = 0 \) and

\[
|f(x)| = \lim_{i \to \infty} |f_i(x)| \leq g(x)
\]

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 \( f_i \in L^1(X \setminus N) \). 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 \( x \in X \setminus N \),

\[
h(x) - g_i(x) = |f(x)| + g(x) - |f_i(x) - f(x)|
\]

\[ \geq |f(x)| + g(x) - (|f_i(x)| + |f(x)|) \]

\[ = g(x) - |f_i(x)| \geq 0. \]
Since \( \lim_{t \to \infty} g_i = 0 \) in \( X \), Fatou’s lemma implies
\[
\int_X h \, d\mu = \int_X \liminf_{i \to \infty} (h - g_i) \, d\mu \\
\leq \liminf_{i \to \infty} \int_X (h - g_i) \, d\mu \\
= \int_X h \, d\mu - \limsup_{i \to \infty} \int_X g_i \, d\mu.
\]
Since \( \int_X h \, d\mu < \infty \), we have
\[
\limsup_{i \to \infty} \int_X g_i \, d\mu \leq 0.
\]
Since \( g_i \geq 0 \), 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
\[
\left| \int_X f_i \, d\mu - \int_X f \, d\mu \right| = \left| \int_X (f_i - f) \, d\mu \right| \leq \int_X |f_i - f| \, d\mu \to 0
\]
as \( i \to \infty \). \( \square \)

**Remarks 3.27:**

(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 \( |f_i - f| \to 0 \) \( \mu \)-almost everywhere and \( |f_i - f| \leq |f_i| + |f| \leq 2g \). 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 \( f_i : X \to [-\infty, \infty] \), \( i = 1, 2, \ldots \), are \( \mu \)-measurable functions such that \( f_i \to f \) \( \mu \)-almost everywhere as \( i \to \infty \). If there exists \( M < \infty \) such that \( |f_i| \leq M \) \( \mu \)-almost everywhere for every \( i = 1, 2, \ldots \), 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 \( \mu(X) < \infty \). \( \blacksquare \)

(4) Assume that \( \mu(X) < \infty \) and \( f_i : X \to [-\infty, \infty] \), \( i = 1, 2, \ldots \), 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 3.28. 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 \( A_i, i = 1, 2, \ldots, \) are pairwise disjoint \( \mu \)-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 \( s_n = \sum_{i=1}^{n} f \chi_{A_i}, n = 1, 2, \ldots, \) 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
\[
|s_n| = \left| \sum_{i=1}^{n} f \chi_{A_i} \right| \leq \sum_{i=1}^{n} |f| \chi_{A_i} \leq \sum_{i=1}^{\infty} |f| \chi_{A_i} \leq |f|,
\]
where \( f \in L^1(X; \mu) \). By the dominated convergence theorem
\[
\int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu = \lim_{n \to \infty} \int_{X} f \chi_{A_i} \, d\mu = \int_{X} \lim_{n \to \infty} s_n \, d\mu
\]
\[
= \lim_{n \to \infty} \int_{X} s_n \, d\mu = \lim_{n \to \infty} \sum_{i=1}^{n} \int_{A_i} f \, d\mu = \sum_{i=1}^{\infty} \int_{A_i} f \, d\mu.
\]
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 \( \mu \)-measurable, \( A_i \supset A_{i+1}, i = 1, 2, \ldots, \) 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 \( |f_i| \leq |f|, f \in L^1(X) \) and \( f_i \to f \chi_A \) everywhere in \( X \) as \( i \to \infty \). By the dominated convergence theorem
\[
\int_{\bigcap_{i=1}^{\infty} A_i} f \, d\mu = \int_{X} f \chi_A \, d\mu = \int_{X} \lim_{i \to \infty} f_i \, d\mu
\]
\[
= \lim_{i \to \infty} \int_{X} f_i \, d\mu = \lim_{i \to \infty} \int_{A_i} f \, d\mu.
\]

(3) (Upwards monotone convergence) If \( A_i, i = 1, 2, \ldots, \) are \( \mu \)-measurable and \( A_i \subset A_{i+1}, i = 1, 2, \ldots, \) then
\[
\int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu = \lim_{i \to \infty} \int_{A_i} f \, d\mu.
\]
CHAPTER 3. INTEGRATION

Reason. Let \( f_i = f \chi_{A_i} \) and denote \( A = \bigcup_{i=1}^{\infty} A_i \). Then \( |f_i| \leq |f|, f \in L^1(X) \) and \( f_i \to f \chi_A \) everywhere in \( X \) as \( i \to \infty \). By the dominated convergence theorem

\[
\int_{\bigcup_{i=1}^{\infty} A_i} f \, d\mu = \int_X f \chi_A \, d\mu = \int_X \lim_{i \to \infty} f_i \, d\mu = \lim_{i \to \infty} \int_X f_i \, d\mu.
\]

Example 3.29. Assume \( f \in L^1(\mathbb{R}^n) \) with respect to the Lebesgue measure. Then

\[
\lim_{i \to \infty} \int_{\mathbb{R}^n} f(x) e^{-\frac{|x|^2}{i}} \, dx = \int_{\mathbb{R}^n} f(x) \, dx.
\]

Reason. Since

\[
|f(x) e^{-\frac{|x|^2}{i}}| \leq |f(x)| \quad \text{for every} \quad x \in \mathbb{R}^n \quad \text{and} \quad i = 1, 2, \ldots,
\]

the function \( f \) itself will do as an integrable majorant in the dominated convergence theorem. Thus

\[
\lim_{i \to \infty} \int_{\mathbb{R}^n} f(x) e^{-\frac{|x|^2}{i}} \, dx = \int_{\mathbb{R}^n} f(x) \lim_{i \to \infty} e^{-\frac{|x|^2}{i}} \, dx = \int_{\mathbb{R}^n} f(x) \, dx.
\]

We conclude this section with two useful results, which are related to integrals depending on a parameter. Assume that \( \mu \) is a measure on \( X \) and let \( I \subset \mathbb{R} \) be an interval. Suppose that for every fixed \( t \in I \) there exists an integrable function on \( X \). Thus we have \( f : X \times I \to [-\infty, \infty], f = f(x, t) \). For each \( t \in I \), we consider the integral of the function over \( X \) and denote

\[
F(t) = \int_X f(x, t) \, d\mu(x).
\]

We are interested in regularity of \( F \). First we discuss continuity.

**Theorem 3.30 (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 \( |f(x, t)| \leq g(x) \) 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

\[
\lim_{t \to t_0} F(t) = \lim_{t \to t_0} \int_X f(x, t) \, d\mu(x)
= \int_X \lim_{t \to t_0} f(x, t) \, d\mu(x)
= \int_X f(x, t_0) \, d\mu(x).
\]
CHAPTER 3. INTEGRATION

**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.

The we discuss differentiability.

**Theorem 3.31 (Differentiability).** Assume that

1. for every $t \in I$, the function $x \rightarrow f(x, t)$ is integrable in $X$,
2. the function $t \rightarrow f(x, t)$ 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 $|\frac{\partial}{\partial t}(x, t)| \leq h(x)$ for every $(x, t) \in X \times I$.

Then $F$ is differentiable at every point $t \in I$ and

$$F'(t) = \frac{\partial}{\partial t} \left( \int_X f(x, t) \mu(x) \right) = \int_X \frac{\partial}{\partial t} f(x, t) d\mu(x).$$

**The Moral:** 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

$$F(t + h) - F(t) = \int_X \frac{f(x, t + h) - f(x, t)}{h} d\mu(x).$$

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

$$\left| \frac{f(x, t + h) - f(x, t)}{h} \right| = \left| \frac{\partial}{\partial t} f(x, t') \right| \leq h(x)$$

for some $t' \in (t, t + h)$. Thus by the dominated convergence theorem

$$F'(t) = \lim_{h \to 0} \int_X \frac{f(x, t + h) - f(x, t)}{h} d\mu(x) = \int_X \frac{\partial}{\partial t} f(x, t) d\mu(x).$$

This kind of arguments are frequently used for Lebesgue integral and partial derivatives in real analysis.
3.7 Lebesgue integral

Lebesgue integrable functions

Let \( f : \mathbb{R}^n \rightarrow [-\infty, \infty] \) be a Lebesgue measurable function. The Lebesgue integral of \( f \) is denoted as

\[
\int_{\mathbb{R}^n} f \, dm = \int_{\mathbb{R}^n} f(x) dm(x) = \int_{\mathbb{R}^n} f(x) dx = \int_{\mathbb{R}^n} f(x) \, dx,
\]

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 3.32. Let \( f : \mathbb{R}^n \rightarrow \mathbb{R}, f(x) = |x|^{-\alpha}, \alpha > 0 \). 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 \( A_i = B(0, 2^{-i+1}) \setminus B(0, 2^{-i}), i = 1, 2, \ldots \). The sets \( A_i \) are Lebesgue measurable, pairwise disjoint and \( B(0, 1) = \bigcup_{i=1}^{\infty} A_i \). Thus

\[\text{Figure 3.1: An exhaustion of } B(0, 1) \text{ by annuli.}\]
\[
\int_{B(0,1)} \frac{1}{|x|^a} \, dx \leq \sum_{i=1}^{\infty} \int_{A_i} \frac{1}{|x|^a} \, dx \leq \sum_{i=1}^{\infty} \int_{A_i} 2^{ia} \, dx = \sum_{i=1}^{\infty} 2^{ia} m(A_i)
\]
\[
\leq \sum_{i=1}^{\infty} 2^{ia} m(B(0,2^{-i+1})) = \sum_{i=1}^{\infty} 2^{ia} 2^{n-i+1} m(B(0,1))
\]
\[
= 2^a m(B(0,1)) \sum_{i=1}^{\infty} 2^{ia-in} < \infty, \quad a < n.
\]

On the other hand,
\[
\int_{B(0,1)} \frac{1}{|x|^a} \, dx = \sum_{i=1}^{\infty} \int_{A_i} \frac{1}{|x|^a} \, dx \geq \sum_{i=1}^{\infty} \int_{A_i} 2^{(i-1)a} \, dx = \sum_{i=1}^{\infty} 2^{(i-1)a} m(A_i)
\]
\[
= (2^a - 1)2^{-a} m(B(0,1)) \sum_{i=1}^{\infty} 2^{ia-in} = \infty, \quad a \geq n.
\]
The last equality follows from
\[
m(A_i) = m(B(0,2^{-i+1})) - m(B(0,2^{-i}))
\]
\[
= (2^{-(i+1)n} - 2^{-in}) m(B(0,1))
\]
\[
= 2^{-in} (2^n - 1) m(B(0,1)).
\]

Thus
\[
\int_{B(0,1)} \frac{1}{|x|^a} \, dx < \infty \iff a < n.
\]
A similar reasoning with the sets \( A_i = B(0,2^i) \setminus B(0,2^{i-1}), \, i = 1, 2, \ldots \), shows that
\[
\int_{\mathbb{R}^n \setminus B(0,1)} \frac{1}{|x|^a} \, dx < \infty \iff a > n.
\]

We shall show in Example 3.36 how to compute these integrals by a change of variables and spherical coordinates.

**Example 3.33.** Let \( \phi : \mathbb{R} \rightarrow [0, \infty) \),
\[
\phi(x) = \begin{cases} 
\frac{1}{\sqrt{x}}, & |x| < 1, \\
0, & |x| \geq 1,
\end{cases}
\]
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
\[
\int_{\mathbb{R}} f(x) \, dx = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} 2^{-i-j} \int_{\mathbb{R}} \phi \left( x - \frac{j}{i} \right) \, dx
\]
\[
= \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} 2^{-i-j} \int_{\mathbb{R}} \phi(x) \, dx
\]
\[
= 4 \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} 2^{-i-j} = 12 < \infty.
\]
Figure 3.2: An exhaustion of $\mathbb{R}^n \setminus B(0,1)$ by annuli.

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 2.45.

$L^1$ 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
\[
\|f\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |f| \, dx = \int_{\mathbb{R}^n} f^+ \, dx + \int_{\mathbb{R}^n} f^- \, dx < \infty.
\]
is called the $L^1$-norm of $f$. This has the usual properties

1. $0 \leq \|f\|_{L^1(\mathbb{R}^n)} < \infty$,
2. $\|f\|_{L^1(\mathbb{R}^n)} = 0 \iff f = 0$ almost everywhere,
(3) \( \| \alpha f \|_{L^1(\mathbb{R}^n)} = |\alpha| \| f \|_{L^1(\mathbb{R}^n)}, \alpha \in \mathbb{R}, \) and
(4) \( \| f + g \|_{L^1(\mathbb{R}^n)} \leq \| f \|_{L^1(\mathbb{R}^n)} + \| g \|_{L^1(\mathbb{R}^n)}. \)

The last triangle inequality in \( L^1 \) follows from the pointwise triangle inequality \( |f(x) + g(x)| \leq |f(x)| + |g(x)|. \) 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.
(4) (Vertical truncation) Approximation by integrals of bounded functions
\[
\int_{\mathbb{R}^n} |f| \, dx = \int_{\mathbb{R}^n} \lim_{i \to \infty} \min \{ |f(0,i)| \} \, dx = \lim_{i \to \infty} \int_{\mathbb{R}^n} \min \{ |f(0,i)| \} \, dx.
\]
Here we again used the monotone convergence theorem or the dominated convergence theorem.

\( L^1 \) 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 3.34:

(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))
\[
m(\{ x \in \mathbb{R}^n : |f_i(x) - f(x)| > \varepsilon \}) \leq \frac{1}{\varepsilon} \int_{\mathbb{R}^n} |f_i(x) - f(x)| \, dx \to 0 \quad \text{as} \quad i \to \infty
\]
for every \( \varepsilon \to 0 \).
Theorem 2.39 implies that if \( f_i \to f \) in \( L^1(\mathbb{R}^n) \), then there exists a subsequence such that \( f_{i_k} \to f \) \( \mu \)-almost everywhere. An example of a sliding sequence of functions, see Example 2.38 (4), shows that the claim is not true for the original sequence.

(2) The Riesz-Fisher theorem states that \( L^1 \) is a Banach space, that is, every Cauchy sequence converges. We shall not prove this in this 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

\[
\int_{\mathbb{R}^n} f(x + x_0) \, dx = \int_{\mathbb{R}^n} \chi_A(x + x_0) \, dx = \int_{\mathbb{R}^n} \chi_{A-x_0}(x) \, dx
\]

\[
= m(A - x_0) = m(A)
\]

\[
= \int_{\mathbb{R}^n} \chi_A(x) \, dx = \int_{\mathbb{R}^n} f(x) \, dx.
\]

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,

\[
\int_{\mathbb{R}^n} f(L^{-1}x) \, dx = |\det L| \int_{\mathbb{R}^n} f(x) \, dx.
\]
CHAPTER 3. INTEGRATION 100

**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

\[
\int_{\mathbb{R}^n} f(Lx) \, dx = \int_{\mathbb{R}^n} \chi_A(Lx) \, dx = \int_{\mathbb{R}^n} (\chi_A \circ L)(x) \, dx = \int_{\mathbb{R}^n} \chi_{L^{-1}(A)}(x) \, dx
\]

\[
= m(L^{-1}(A)) = |\det L^{-1}| m(A)
\]

\[
= \frac{1}{|\det L|} \int_{\mathbb{R}^n} \chi_A(x) \, dx = \frac{1}{|\det L|} \int_{\mathbb{R}^n} f(x) \, dx.
\]

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 [4] pages 170–171 and 65–80 or or [12] 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.

**Reason.**

\[
m(L(A)) = \int_{L(A)} 1 \, dx = \int_{\mathbb{R}^n} \chi_{L(A)}(x) \, dx = \int_{\mathbb{R}^n} (\chi_A \circ L^{-1})(x) \, dx = \int_{\mathbb{R}^n} \chi_A(L^{-1}(x)) \, dx = |\det L| \int_{\mathbb{R}^n} \chi_A(x) \, dx = |\det L| m(A).
\]

Moreover, \( A \) is a Borel set if and only if \( L(A) \) is a Borel set, since \( L \) is a homeomorphism.

(5) (Nonlinear change of variables) Let \( U \subset \mathbb{R}^n \) be an open set and suppose that \( \Phi: U \to \mathbb{R}^n \), \( \Phi = (\phi_1, \ldots, \phi_n) \) is a \( C^1 \) diffeomorphism. We denote by \( D\Phi \) the derivative matrix with entries \( D\phi_i \), \( i, j = 1, \ldots, n \). The mapping \( \Phi \) 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 \( \phi_i \), \( i = 1, \ldots, n \) 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 \( \Phi(U) \), the \( f \circ \Phi \) 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 \( \Phi(U) \) 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 [4] pages 494–503 or [12] pages 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 $\Phi$ is a linear mapping, then $D\Phi = \Phi$ and this is compatible with the change of variables formula for linear mappings.

Example 3.35. 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

$$U = (0, \infty) \times (0, \pi)^{n-2} \times (0, 2\pi) \subset \mathbb{R}^n, \quad n \geq 2.$$

Denote the coordinates of a point in $U$ by $r, \theta_1, \ldots, \theta_{n-2}, \theta_{n-1}$, respectively. We define $\Phi: U \to \mathbb{R}^n$ by the spherical coordinate formulas as follows. If $x = \Phi(r, \theta)$, then

$$x_i = r \sin \theta_1 \cdots \sin \theta_{i-1} \cos \theta_i, \quad i = 1, \ldots, n,$$

where $\theta_n = 0$ so that $x_n = r \sin \theta_1 \cdots \sin \theta_{n-1}$. Then $\phi$ 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

$$\int_{\mathbb{R}^n} f(x) \, dx = \int_0^\infty \ldots \int_0^\infty f(\Phi(r, \theta)) r^{n-1} (\sin \theta_1)^{n-2} \ldots (\sin \theta_{n-3})^2 \sin \theta_{n-2} \, d\theta_{n-1} \ldots d\theta_1 \, dr.$$  

It can be shown that

$$\omega_{n-1} = \int_0^\pi (\sin \theta_1)^{n-2} \, d\theta_{n-1} \ldots \int_0^\pi (\sin \theta_{n-3})^2 \, d\theta_{n-3} \int_0^\pi \sin \theta_{n-2} \, d\theta_{n-2} \int_0^{2\pi} \, d\theta_{n-1},$$

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

$$\int_{\mathbb{R}^n} f(|x|) \, dx = \omega_{n-1} \int_0^\infty f(r) r^{n-1} \, dr.$$  

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

\[
\begin{align*}
r^n \Omega_n &= r^n m(B(0, 1)) = m(B(x, r)) = m(B(0, r)) \\
&= \int_{\mathbb{R}^n} \chi_{B(0,r)}(y) dy = \int_{\mathbb{R}^n} \chi_{(0,r)}(|y|) dy \\
&= \omega_{n-1} \int_0^r \rho^{n-1} d\rho = \omega_{n-1} \frac{r^n}{n}.
\end{align*}
\]

In particular, it follows that \( \omega_{n-1} = n \Omega_n \) and

\[
m(B(x, r)) = \frac{2\pi^{n/2}}{\Gamma(n/2)} \frac{r^n}{n} = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)} r^n.
\]

**Example 3.36.** Let \( r > 0 \). Then by property (3) above and Example 3.32,

\[
\begin{align*}
\int_{\mathbb{R}^n \setminus B(0,r)} \frac{1}{|x|^a} dx &= \int_{\mathbb{R}^n} \frac{1}{|x|^a} \chi_{\mathbb{R}^n \setminus B(0,r)}(x) dx \\
&= r^n \int_{\mathbb{R}^n} \frac{1}{|rx|^a} \chi_{\mathbb{R}^n \setminus B(0,r)}(rx) dx \\
&= r^{n-a} \int_{\mathbb{R}^n} \frac{1}{|x|^a} \chi_{\mathbb{R}^n \setminus B(0,1)}(x) dx \\
&= r^{n-a} \int_{\mathbb{R}^n \setminus B(0,1)} \frac{1}{|x|^a} dx < \infty, \quad a > n,
\end{align*}
\]

and, in a similar way,

\[
\int_{B(0,r)} \frac{1}{|x|^a} dx = r^{n-a} \int_{B(0,1)} \frac{1}{|x|^a} dx < \infty, \quad a < n.
\]

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

\[
\begin{align*}
\int_{\mathbb{R}^n \setminus B(0,r)} \frac{1}{|x|^a} dx &= \omega_{n-1} \int_r^\infty \rho^{-a} \rho^{n-1} d\rho \\
&= \omega_{n-1} \frac{\rho^{-a+n}}{-a+n} \bigg|_r^\infty = \omega_{n-1} \frac{\rho^{-a+n}}{-a+n} < \infty, \quad a > n
\end{align*}
\]

and

\[
\begin{align*}
\int_{B(0,r)} \frac{1}{|x|^a} dx &= \omega_{n-1} \int_0^r \rho^{-a} \rho^{n-1} d\rho \\
&= \omega_{n-1} \frac{\rho^{-a+n}}{-a+n} \bigg|_0^r = \omega_{n-1} \frac{\rho^{-a+n}}{-a+n} r^{-a} < \infty, \quad a < n.
\end{align*}
\]

**Remarks 3.37:**

Formula (3.1) (or Example 3.32) implies following claims:

1. If \(|f(x)| < c|x|^{-a}\) in a ball \( B(0, r) \), \( r > 0 \), for some \( a < n \), then \( f \in L^1(B(0, r)) \).

On the other hand, if \(|f(x)| > c|x|^{-a}\) in \( B(0, r) \) for some \( a > n \), then \( f \notin L^1(B(0, r)) \).
(2) If $|f(x)| \leq c|x|^{-\alpha}$ 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 $|f(x)| \geq c|x|^{-\alpha}$ 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 3.38.** 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:**

1. Since $f = f^+ - f^-$, we may assume that $f \geq 0$. Then there is an increasing sequence of simple functions $f_i$, $i = 1, 2, \ldots$, such that $f_i \to f$ everywhere in $\mathbb{R}^n$ as $i \to \infty$. By the dominated convergence theorem, we have

$$
\lim_{i \to \infty} \|f_i - f\|_{L^1(\mathbb{R}^n)} = \lim_{i \to \infty} \int_{\mathbb{R}^n} |f_i - f| \, dx = \int_{\mathbb{R}^n} \lim_{i \to \infty} |f_i - f| \, dx = 0,
$$

because $|f_i - f| \leq |f_i| + |f| \leq 2|f| \in L^1(\mathbb{R}^n)$ gives an integrable majorant.

2. **Step 1:** Since $f = f^+ - f^-$ we may assume that $f \geq 0$.

**Step 2:** The dominated convergence theorem gives

$$
\lim_{i \to \infty} \|f \chi_{B(0, i)} - f\|_{L^1(\mathbb{R}^n)} = \lim_{i \to \infty} \int_{\mathbb{R}^n} |f \chi_{B(0, i)} - f| \, dx = \int_{\mathbb{R}^n} \lim_{i \to \infty} |f \chi_{B(0, i)} - f| \, dx = 0,
$$

since $|f \chi_{B(0, i)} - f| \leq |f| \in L^1(\mathbb{R}^n)$. Thus compactly supported integrable functions are dense in $L^1(\mathbb{R}^n)$.

**Step 3:** There is an increasing sequence of simple functions $f_i : \mathbb{R}^n \to [0, \infty)$, $i = 1, 2, \ldots$, such that $f_i \to f$ everywhere in $\mathbb{R}^n$ as $i \to \infty$. The dominated convergence theorem gives

$$
\lim_{i \to \infty} \|f_i - f\|_{L^1(\mathbb{R}^n)} = \lim_{i \to \infty} \int_{\mathbb{R}^n} |f_i - f| \, dx = \int_{\mathbb{R}^n} \lim_{i \to \infty} |f_i - f| \, dx = 0.
$$

since $|f_i - f| \leq |f_i| + |f| \leq 2|f| \in L^1(\mathbb{R}^n)$. 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 \( a_i > 0, i = 1, 2, \ldots \). Thus if we can approximate each \( \chi_{A_i} \) 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 1.37 there exist a compact set \( K \) and an 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. \( 0 \leq g(x) \leq 1 \) 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

\[
U = \{ x \in \mathbb{R}^n : \text{dist}(x, K) < \frac{1}{2} \text{dist}(K, \mathbb{R}^n \setminus G) \}.
\]

Then \( K \subset U \subset G, U \) is open and \( \overline{U} \) is compact. The function \( g : \mathbb{R}^n \to \mathbb{R} \),

\[
g(x) = \frac{\text{dist}(x, \mathbb{R}^n \setminus U)}{\text{dist}(x, K) + \text{dist}(x, \mathbb{R}^n \setminus U)},
\]

has the desired properties, see Remark 1.22. Moreover,

\[
\text{supp } g = \{ x \in \mathbb{R}^n : g(x) \neq 0 \} = \overline{U}
\]

is compact. \( \blacksquare \)

Observe that

\[
x \in K \implies \chi_{A}(x) - g(x) = 1 - 1 = 0,
\]

\[
x \in \mathbb{R}^n \setminus G \implies \chi_{A}(x) - g(x) = 0 - 0 = 0,
\]

\[
x \in A \setminus K \implies \chi_{A}(x) - g(x) = 1 - g(x) < 1,
\]

\[
x \in G \setminus A \implies \chi_{A}(x) - g(x) = g(x) < 1.
\]

Thus \(|\chi_{A} - g| \leq 1, \chi_{A} - g \) vanishes in \( K \) and outside \( G \setminus K \) and we have

\[
\|f - g\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n} |\chi_{A} - g| \, dx \leq m(G \setminus K) < \varepsilon.
\]

This completes the proof of the approximation property. \( \square \)

**Remark 3.39.** The claim (2) can also be proved using Lusin’s theorem (Theorem 2.45)(exercise).
3.8 Cavalieri’s principle

Recall, that every nonnegative measurable function satisfies Tchebyshev’s inequality (Theorem 3.8 (5))

\[ m(\{ x \in \mathbb{R}^n : f(x) > t \}) \leq \frac{1}{t} \int_{\mathbb{R}^n} f \, dx, \quad t > 0. \]

In particular, if \( f \in L^1(\mathbb{R}^n) \), then

\[ m(\{ x \in \mathbb{R}^n : f(x) > t \}) \leq \frac{c}{t}, \quad t > 0. \] (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 \( tm(\{ x \in \mathbb{R}^n : f(x) > t \}) \leq c \) for every \( t > 0 \) if \( f \in L^1(\mathbb{R}^n) \), but there is a stronger result.

Lemma 3.40. 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

\[
\lim_{t \to \infty} \int_{\mathbb{R}^n \setminus A_t} f \, dx = \lim_{t \to \infty} \int_{\mathbb{R}^n} \chi_{\mathbb{R}^n \setminus A_t} f \, dx \\
= \int_{\mathbb{R}^n} \lim_{t \to \infty} f \, dx = \int_{\mathbb{R}^n} \chi_A f \, dx \\
= \int_A f \, dx = \int_A f \, dx + \int_{\mathbb{R}^n \setminus A} f \, dx = \int_{\mathbb{R}^n} f \, dx.
\]
Thus
\[
\int_{\mathbb{R}^n} f \, dx = \lim_{t \to \infty} \left( \int_{A_t} f \, dx + \int_{\mathbb{R}^n \setminus A_t} f \, dx \right)
\]
\[
= \lim_{t \to \infty} \int_{A_t} f \, dx + \int_{\mathbb{R}^n} f \, dx,
\]
which implies that
\[
\lim_{t \to \infty} \int_{A_t} f \, dx = 0.
\]
By Tchebyshev's inequality
\[
0 \leq t m(A_t) \leq \int_{A_t} f \, dx \to 0 \quad \text{as} \quad t \to \infty,
\]
which implies \(t m(A_t) \to 0\) as \(t \to \infty\). □

The following representation of the integral is very useful.

**Theorem 3.41 (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
\[
\int_A f \, dx = \int_0^{\infty} m(\{x \in A : f(x) > t\}) dt.
\]

**The moral:** In order to estimate the integral of a function it is enough to estimate the distribution sets of the function.

**Remarks 3.42:**

(1) For signed Lebesgue measurable functions we have
\[
\int_A |f| \, dx = \int_0^{\infty} m(\{x \in A : |f(x)| > t\}) dt.
\]
This could be taken as the definition of the integral.

(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 \(t > 0\), then \(\int_A g \, dx = \int_A f \, dx\).

(4) 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 3.46.
Figure 3.5: Cavelieri’s principle.

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

\[
\int_{0}^{\infty} m(\{x \in A : f(x) > t\}) \, dt = \int_{0}^{a_k} m(\{x \in A : f(x) > t\}) \, dt
\]

\[
= \sum_{i=1}^{k} \int_{a_{i-1}}^{a_i} m(\{x \in A : f(x) > t\}) \, dt
\]

\[
= \sum_{i=1}^{k} (a_i - a_{i-1}) m \left( \bigcup_{j=i}^{k} A_j \cap A \right)
\]

\[
= \sum_{i=1}^{k} (a_i - a_{i-1}) \sum_{j=i}^{k} m(A_j \cap A) \quad (A_j \text{ measurable and disjoint})
\]

\[
= \sum_{i=1}^{k} a_i \sum_{j=i}^{k} m(A_j \cap A) - \sum_{i=1}^{k} a_{i-1} \sum_{j=i}^{k} m(A_j \cap A)
\]

\[
= \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=1}^{j} a_i - \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=0}^{j-1} a_i
\]

\[
= \sum_{j=1}^{k} m(A_j \cap A) \sum_{i=1}^{j} (a_i - a_{i-1}) = \sum_{j=1}^{k} a_j m(A_j \cap A) = \int_{A} f \, dx.
\]

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
Figure 3.6: Cavalieri’s principle for a simple function.

\( f_i \leq f_{i+1} \) 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 \( \varphi_j \) is an increasing sequence of functions and \( \varphi_i(t) \to \varphi(t) \) for every \( t \geq 0 \) as \( i \to \infty \). The monotone convergence theorem implies

\[
\int_0^{\infty} m(\{x \in A : f(x) > t\}) dt = \lim_{i \to \infty} \int_0^{\infty} m(\{x \in A : f_i(x) > t\}) dt
\]

\[ = \lim_{i \to \infty} \int_A f_i \, dx = \int_A f \, dx. \]

Remarks 3.43:

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 3.58.

Example 3.44. Let \( f : \mathbb{R}^n \to \mathbb{R}, f(x) = |x|^{-\alpha}, 0 < \alpha < n \). Then
\[
\int_{B(0,1)} f(x) \, dx = \int_{B(0,1)} |x|^{-\alpha} \, dx = \int_0^\infty m(\{ x \in B(0,1) : |x|^{-\alpha} > t \}) \, dt
\]
\[
= \int_0^1 m(B(0,1)) \, dt + \int_1^\infty m(\{ x \in \mathbb{R}^n : |x| < t^{-1/\alpha} \}) \, dt
\]
\[
= m(B(0,1)) + \int_1^\infty t^{-n/\alpha} m(B(0,1)) \, dt
\]
\[
= m(B(0,1)) \left( 1 + \frac{\alpha}{n - \alpha} \right).
\]

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 \( I_i, i = 1, \ldots, k \), be pairwise disjoint intervals in \( \mathbb{R} \) with \( \bigcup_{i=1}^k I_i = [a, b] \) with \( a, b \in \mathbb{R} \) and let \( a_i, i = 1, \ldots, k \), be real numbers. A function \( f : [a, b] \to \mathbb{R} \) is said to a step function, if
\[
f = \sum_{i=1}^k a_i \chi_{I_i}.
\]
Observe, that a step function is just a special type simple function. Let \( f : [a, b] \to \mathbb{R} \) be a bounded function. Recall that the lower Riemann integral is
\[
\int_a^b f(x) \, dx = \sup \left\{ \int_{[a,b]} g \, dx : g \leq f \text{ on } [a,b] \text{ and } g \text{ is a step function} \right\}
\]
and the upper Riemann integral is
\[
\int_a^b f(x) \, dx = \inf \left\{ \int_{[a,b]} h \, dx : f \leq h \text{ on } [a,b] \text{ and } h \text{ is a step function} \right\}.
\]
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 \( f : [a, b] \rightarrow \mathbb{R} \) is a bounded Lebesgue measurable function, then by the definition of the Lebesgue integral,
\[
\int_a^b f(x) \, dx = \int_{[a, b]} f(x) \, dx = \int_a^b f(x) \, dx.
\]
This implies that if \( f \) is Riemann integrable, then the Riemann and Lebesgue integrals coincide provided \( f \) is Lebesgue measurable.

**Lemma 3.45.** Let \( f : [a, b] \rightarrow \mathbb{R} \) 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 \( f \geq 0 \). By the definition of the Riemann integral, there are step functions \( g_i \) and \( h_i \) such that \( g_i \leq f \leq h_i \),
\[
\lim_{i \to \infty} \int_{[a, b]} g_i \, dx = \lim_{i \to \infty} \int_{[a, b]} h_i \, dx = \int_a^b f(x) \, dx.
\]
By passing to sequences max{\( g_1, \ldots, g_j \)} and min{\( h_1, \ldots, h_j \)} we may assume that \( g_i \leq g_{i+1} \) and \( h_i \geq h_{i+1} \) for every \( i = 1, 2, \ldots \). 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 \( h - g \geq 0 \) and
\[
\int_{[a, b]} (h(x) - g(x)) \, dx = \int_{[a, b]} h(x) \, dx - \int_{[a, b]} g(x) \, dx
\]
\[
= \int_a^b f(x) \, dx - \int_a^b f(x) \, dx = 0,
\]
we have \( h - g = 0 \) almost everywhere in \([a, b]\). Since \( g \leq f \leq h \) 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 3.46.** 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 things get 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 \( f_i \in L^1([a, b]) \), \( i = 1, 2, \ldots \), are functions such that

\[
\sum_{i=1}^\infty \| f_i \|_{L^1([a, b])} < \infty,
\]

then Corollary 3.15 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 3.47:**

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 \( q_i, i = 1, 2, \ldots, \) be an enumeration of rational numbers in the interval \([0,1]\) and define \( f_i : \mathbb{R} \to \mathbb{R}, \)

\[
f_i(x) = \chi_{[q_1, \ldots, q_i]}(x), \quad i = 1, 2, \ldots
\]

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.
(3) Define \( f : [0, 1] \to \mathbb{R} \) by setting \( f(0) = 0 \) and

\[
f(x) = \begin{cases} \frac{2^i + 1}{i}, & \frac{1}{2^i} < x \leq \frac{3}{2^{i+1}}, \\ \frac{2^i - 1}{i}, & \frac{3}{2^{i+1}} < x \leq \frac{1}{2^{i+1}}, \end{cases}
\]

for \( x \in (0, 1] \). Note that \( \frac{3}{2^{i+1}} \) is the midpoint of the interval \([\frac{1}{2^i}, \frac{1}{2^{i+1}}]\) 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:** \( f \not\in L^1([0, \infty)) \).

**Reason.**

\[
\int_{[0,\infty)} |f| \, dx = \int_{0}^{\pi} |f(x)| \, dx + \sum_{i=1}^{\infty} \int_{i \pi}^{(i+1) \pi} \frac{|\sin x|}{|x|} \, dx
\]

\[
\geq \int_{0}^{\pi} |f(x)| \, dx + \sum_{i=1}^{\infty} \frac{1}{i+1} = \infty.
\]

**Claim:** The improper Riemann integral \( \lim_{a \to \infty} \int_{0}^{a} \frac{\sin x}{x} \, dx \) exists.

**Reason.** Denote \( I(a) = \int_{0}^{a} \frac{\sin x}{x} \, dx, \ a \geq \pi \). Then

\[
I(k\pi) = \int_{0}^{k\pi} \frac{\sin x}{x} \, dx = \sum_{i=0}^{k-1} \int_{i \pi}^{(i+1) \pi} \frac{\sin x}{x} \, dx, \ k = 1, 2, \ldots,
\]

where

\[
\sum_{i=0}^{k-1} a_i \, dx \quad \text{with} \quad a_i = \int_{i \pi}^{(i+1) \pi} \frac{\sin x}{x} \, dx, \ i = 1, 2, \ldots,
\]

is an alternating series with the properties

\[
a_i a_{i+1} < 0, \quad |a_i| \leq |a_{i+1}| \quad \text{and} \quad \lim_{i \to \infty} a_i = 0.
\]
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 \( a \geq \pi \), we have \( a \in [k\pi, (k+1)\pi) \) for some \( k = 1, 2, \ldots \), and
\[ |I(a) - I(k\pi)| = \left| \int_{k\pi}^{a} \frac{\sin x}{x} \, dx \right| \leq \int_{k\pi}^{a} \frac{1}{k} \, dx \leq \frac{1}{k}. \]

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 3.48.** 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 \( \sum_{i=1}^{\infty} \text{vol}(I_i) = \frac{1}{2} \). For every \( i = 1, 2, \ldots \), choose a closed centered subinterval \( J_i \subset I_i \) such that \( \text{vol}(J_i) = \text{vol}(I_i)^2 \).

Define a continuous function \( f : [0, 1] \to \mathbb{R} \) such that
\[ f(x) = 0 \quad \text{for every} \quad x \in [0, 1] \setminus \bigcup_{i=1}^{\infty} J_i, \]
\[ 0 \leq f(x) \leq 1 \quad \text{for every} \quad x \in [0, 1] \] 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 \( x \in [0, 1] \) (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 3.49.** Let \( \mu^* \) be an outer measure on \( X \) and \( \nu^* \) 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 $\mu^*$-measurable sets $A_i \subset X$ and $\nu^*$-measurable sets $B_i \subset Y$, $i = 1, 2, \ldots$. The measure $\mu^* \times \nu^*$ is called the product measure of $\mu^*$ and $\nu^*$.

**Remark 3.50.** It is an exercise to show that $\mu^* \times \nu^*$ is an outer measure on $X \times Y$.

**Remark 3.50.** 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 $\mu^*$-measurable set and $B \subset Y$ is a $\nu^*$-measurable set, see Fubini’s theorem below.

**Theorem 3.51 (Fubini’s theorem).** Let $\mu^*$ be an outer measure on $X$ and $\nu^*$ an outer measure on $Y$.

1. Then $\mu^* \times \nu^*$ is a regular outer measure on $X \times Y$, even if $\mu^*$ and $\nu^*$ are not regular.
2. If $A \subset X$ is a $\mu^*$-measurable set and $B \subset Y$ is a $\nu^*$-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 $\sigma$-finite with respect to $\mu^* \times \nu^*$, then $S_y = \{x \in X : (x, y) \in S\}$ is $\mu^*$-measurable for $\nu^*$-almost every $y \in Y$ and $S_x = \{y \in Y : (x, y) \in S\}$ is $\nu^*$-measurable for $\nu^*$-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 and $|f| \neq 0$ is $\sigma$-finite with respect to $\mu^* \times \nu^*$, 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 $\nu^*$-measurable function,
   $$x \mapsto \int_Y f(x, y) d\nu(y)$$
   is a $\mu^*$-measurable function and
   $$\int_{X \times Y} f(x, y) d(\mu \times \nu) = \int_Y \left[ \int_X f(x, y) d\mu(x) \right] d\nu(y)$$
   $$= \int_X \left[ \int_Y f(x, y) d\nu(y) \right] d\mu(x).$$

**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. Let $\mathcal{F}$ 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 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 $S_i \in \mathcal{F}$, $i = 1, 2, \ldots$, are pairwise disjoint, then $S = \bigcup_{i=1}^{\infty} S_i \in \mathcal{F}$.

Reason. Note that $\chi_S = \sum_{i=1}^{\infty} \chi_{S_i}$. By Corollary 3.15 we have $\rho(S) = \sum_{i=1}^{\infty} \rho(S_i)$. This shows that $\mathcal{F}$ is closed under countable unions of pairwise disjoint sets. ■

Claim: If $S_i \in \mathcal{F}$, $i = 1, 2, \ldots$, $S_1 \supset S_2 \supset \ldots$, and $\rho(S_1) < \infty$, then $S = \bigcap_{i=1}^{\infty} S_i \in \mathcal{F}$.

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 $\mathcal{F}$ is closed under decreasing convergence of sets with a finiteness condition. ■

Define
\[
\mathcal{P}_0 = \{A \times B : A \text{ is } \mu^*-\text{measurable and } B \text{ is } \nu^*-\text{measurable}\},
\]
\[
\mathcal{P}_1 = \left\{ \bigcup_{i=1}^{\infty} S_i : S_i \in \mathcal{P}_0 \right\} \quad \text{and} \quad \mathcal{P}_2 = \left\{ \bigcap_{i=1}^{\infty} S_i : S_i \in \mathcal{P}_0 \right\}.
\]
The members of $\mathcal{P}_0$ are called measurable rectangles, the class $\mathcal{P}_1$ consists of countable unions of measurable rectangles and and $\mathcal{P}_2$ of countable intersections of these. The latter sets constitute a class relative to which the product measure will be regular.

Note that $\mathcal{P}_0 \subset \mathcal{F}$ and
\[ \rho(A \times B) = \mu(A)\nu(B) \]
whenever $A \times B \in \mathcal{P}_0$. If $A_1 \times B_1, A_2 \times B_2 \in \mathcal{P}_0$, then
\[ (A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2) \in \mathcal{P}_0 \]
and
\[ (A_1 \times B_1) \setminus (A_2 \times B_2) = ((A_1 \setminus A_2) \times B_1) \cup ((A_1 \cap A_2) \times (B_1 \setminus B_2)) \]
is a disjoint union of members of $\mathcal{P}_0$. It follows that every member of $\mathcal{P}_1$ is a countable union of pairwise disjoint members of $\mathcal{P}_0$ and hence $\mathcal{P}_1 \subset \mathcal{F}$.

Claim: For every $S \subset X \times Y$,
\[ (\mu^* \times \nu^*) (S) = \inf \{ \rho(R) : S \subset R \in \mathcal{P}_1 \} \].
\section*{CHAPTER 3. INTEGRATION}

\textbf{Reason.} Suppose that $A_i \times B_i \in \mathcal{P}_0$, $i = 1, 2, \ldots$ and $S \subset R = \bigcup_{i=1}^{\infty} (A_i \times B_i)$. Then $\chi_R = \sum_{i=1}^{\infty} \chi_{A_i \times B_i}$ and

$$
\rho(R) \leq \sum_{i=1}^{\infty} \rho(A_i \times B_i) = \sum_{i=1}^{\infty} \mu^*(A_i) v^*(B_i).
$$

Thus

$$
\inf \{ \rho(R) : S \subset R \in \mathcal{P}_1 \} \leq (\mu^* \times v^*)(S).
$$

Moreover, if $R = \bigcup_{i=1}^{\infty} (A_i \times B_i)$ is any such set, there exist pairwise disjoint sets $A'_i \times B'_i \in \mathcal{P}_0$ such that

$$
R = \bigcup_{i=1}^{\infty} (A_i \times B_i) = \bigcup_{i=1}^{\infty} (A'_i \times B'_i).
$$

Thus

$$
\rho(R) = \sum_{i=1}^{\infty} \mu^*(A'_i) v^*(B'_i) \geq (\mu^* \times v^*)(S).
$$

This shows that the equality holds. \hfill \blacksquare

\textbf{Fix} $A \times B \in \mathcal{P}_0$. Then

$$(\mu^* \times v^*)(A \times B) \leq \mu^*(A) v^*(B) = \rho(A \times B) \leq \rho(R)$$

for every $R \in \mathcal{P}_1$ such that $A \times B \subset R$. Thus the claim above implies

$$(\mu^* \times v^*)(A \times B) = \mu^*(A) v^*(B).$$

We show that $A \times B$ is $\mu^* \times v^*$-measurable. Let $R \in \mathcal{P}_1$ with $T \subset R$. Then $R \setminus (A \times B)$ and $R \cap (A \times B)$ are disjoint members of $\mathcal{P}_1$. Thus

$$
(\mu^* \times v^*)(T \setminus (A \times B)) + (\mu^* \times v^*)(T \cap (A \times B)) \leq \rho(R \setminus (A \times B)) + \rho(R \cap (A \times B)) = \rho(R).
$$

The claim above implies

$$(\mu^* \times v^*)(T \setminus (A \times B)) + (\mu^* \times v^*)(T \cap (A \times B)) \leq (\mu^* \times v^*)(T).$$

Since this holds for all $T \subset X \times Y$, we have shown that $A \times B$ is $\mu^* \times v^*$-measurable. Moreover, it follows that $\mathcal{P}_0$ and $\mathcal{P}_1$ consist of $\mu^* \times v^*$-measurable sets. This proves claim (2) of the theorem.

\textbf{Next we show that $\mu^* \times v^*$ is a regular measure.}

\textbf{Claim:} For every $S \subset X \times Y$ there is $R \in \mathcal{P}_2$ such that $S \subset R$ and

$$(\mu^* \times v^*)(S) = (\mu^* \times v^*)(R) = \rho(R).$$
Claim (4) reduces to (3) when

We write $f$ with $X$.

This proves claim (3) of the theorem, because the other formula is symmetric.

It follows that $\mu^* \times \nu^*$ is $\mathcal{P}_2$-regular. The claim follows from this, since every set in $\mathcal{P}_2$ is $\mu^* \times \nu^*$-measurable by claim 2 of the theorem.

If $S \subseteq X \times Y$ with $(\mu^* \times \nu^*)(S) = 0$, then there exists a set $R \in \mathcal{P}_2$ such that $S \subseteq R$ and $\rho(R) = 0$. Thus $S \in \mathcal{F}$ and $\rho(S) = 0$.

Suppose that $S \subseteq X \times Y$ is $\mu^* \times \nu^*$-measurable and $(\mu^* \times \nu^*)(S) < \infty$. Then there is $R \in \mathcal{P}_2$ such that $S \subseteq R$ 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 $\nu^*$-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 $\mu^*$ by $\nu^*$. The extension to $\sigma$-finite $\mu^* \times \nu^*$-measure is obvious.

Claim (4) reduces to (3) when $f = \chi_S$. If $f$ is a nonnegative $\mu^* \times \nu^*$-measurable function and is $\sigma$-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 3.52 (Tonelli’s theorem).** Let $\mu^*$ be an outer measure on $X$ and $\nu^*$ an outer measure on $Y$ and suppose that both measures are $\sigma$-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 $\nu^*$-measurable function,

$$x \mapsto \int_X f(x,y) d\nu(y)$$
is a $\mu^*$-measurable function and
\[
\int_{X \times Y} f(x,y) d(\mu \times \nu) = \int_Y \left[ \int_X f(x,y) d\mu(x) \right] d\nu(y) \\
= \int_X \left[ \int_Y f(x,y) d\nu(y) \right] d\mu(x).
\]

**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 3.53:**

1. If $\mu^*$ and $\nu^*$ are counting measures, Tonelli’s theorem reduces to a corresponding claim for series. Let $x_{i,j} \in [0, \infty]$, $i, j = 1, 2, \ldots$. 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.

2. Let $\mu^*$ be an outer measure on $X$ and $\nu^*$ an outer measure on $Y$ and suppose that both measures are $\sigma$-finite. Let $f : X \to [-\infty, \infty]$ be a $\mu^* \times \nu^*$-measurable function. If any of the three integrals
   \[
   \int_{X \times Y} |f(x,y)| d(\mu \times \nu), \quad \int_Y \left[ \int_X |f(x,y)| d\mu(x) \right] d\nu(y), \quad \int_X \left[ \int_Y |f(x,y)| d\nu(y) \right] d\mu(x)
   \]
   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 $\nu^*$-integrable for $\mu^*$-almost every $x \in X$ and that the function $x \mapsto f(x,y)$ is $\mu^*$-integrable for $\nu^*$-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. 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
\[
m^{n+m}(A \times B) = m^s(A)m^m(B),
\]
with the understanding that if one of the sets is of measure zero, then the product set is of measure zero.
Theorem 3.54 (Tonelli’s theorem). Let \( f : \mathbb{R}^{n+m} \to [0, \infty] \) be a nonnegative \( m^{n+m} \)-measurable function. Then
\[
y \mapsto \int_X f(x, y) \, dm^n(x)
\]
is a \( m^m \)-measurable function,
\[
x \mapsto \int_X f(x, y) \, dm^m(y)
\]
is a \( m^n \)-measurable function and
\[
\int_{\mathbb{R}^{n+m}} f(x, y) \, dm^{n+m} = \int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^m} f(x, y) \, dm^m(y) \right] \, dm^n(x) = \int_{\mathbb{R}^m} \left[ \int_{\mathbb{R}^n} f(x, y) \, dm^n(x) \right] \, dm^m(y).
\]

Theorem 3.55 (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
\[
\int_{\mathbb{R}^{n+m}} |f(x, y)| \, dm^{n+m},
\]
\[
\int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^m} |f(x, y)| \, dm^m(y) \right] \, dm^n(x),
\]
\[
\int_{\mathbb{R}^m} \left[ \int_{\mathbb{R}^n} |f(x, y)| \, dm^n(x) \right] \, dm^m(y),
\]
is finite.

1. If \( A \subset \mathbb{R}^n \) is a \( \mu^n \)-measurable set and \( B \subset \mathbb{R}^m \) is a \( m^m \)-measurable set, then
\( A \times B \) is \( m^{n+m} \) measurable and \( (m^{n+m})(A \times B) = m^n(A)m^m(B) \).

2. \( y \mapsto \int_X f(x, y) \, dm^n(x) \)
is a \( m^m \)-integrable function
\[
x \mapsto \int_X f(x, y) \, dm^m(y)
\]
is a \( m^n \)-integrable function and
\[
\int_{\mathbb{R}^{n+m}} f(x, y) \, dm^{n+m} = \int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^m} f(x, y) \, dm^m(y) \right] \, dm^n(x) = \int_{\mathbb{R}^m} \left[ \int_{\mathbb{R}^n} f(x, y) \, dm^n(x) \right] \, dm^m(y).
\]

We consider few corollaries of the previous theorems.

Corollary 3.56. Suppose that \( f : \mathbb{R}^n \to [-\infty, \infty] \) is a \( m^n \)-measurable function. Then the function \( \tilde{f} : \mathbb{R}^m \to [-\infty, \infty] \) defined by \( \tilde{f}(x, y) = f(x) \) 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
\[(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : \tilde{f}(x, y) < a = A \times \mathbb{R}^m,
\]
we conclude that the set is $m^{n+m}$-measurable for every $a \in \mathbb{R}$. Thus $\tilde{f}$ is a $m^{n+m}$-measurable function. \qed

Corollary 3.57. Assume that $f : \mathbb{R}^n \to [0, \infty]$ is a nonnegative function and let
\[A = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : 0 \leq y \leq f(x)\}.
\]

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
\[
\int_{\mathbb{R}^n} f \, dm^n = m^{n+1}(A).
\]

THEOREM: The integral describes the area under the graph of a function.

Proof. Assume that $f$ is a $m^n$-measurable function. Then the function
\[F(x, y) = y - f(x)
\]
is $m^{n+1}$-measurable and thus
\[A = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y \geq 0\} \cap \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : F(x, y) \leq 0\}
\]
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
\[
m^{n+1}(A) = \int_{\mathbb{R}^{n+1}} \chi_A(x, y) \, dm^{n+1}(x, y)
= \int_{\mathbb{R}^n} m(A_x) \, dm^n(x)
= \int_{\mathbb{R}^n} f(x) \, dm^n(x).
\]

We give an alternative proof for Cavalieri's principle by Fubini's theorem. Compare to Theorem 3.41.

Corollary 3.58. 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.

\[
\int_A |f| \, dx = \int_{\mathbb{R}^n} \chi_A(x)f(x) \, dx
\]

\[
= \int_{\mathbb{R}^n} \int_0^{\infty} \chi_A(x)\chi_{[0,f(x))}(t) \, dt \, dx
\]

\[
= \int_0^{\infty} \int_{\mathbb{R}^n} \chi_A(x)\chi_{[0,f(x))}(t) \, dx \, dt \quad \text{(Fubini)}
\]

\[
= \int_0^{\infty} \int_{\mathbb{R}^n} \chi_A(x)\chi_{\{x \in \mathbb{R}^n : f(x) > t\}}(x) \, dx \, dt
\]

\[
= \int_0^{\infty} m(\{x \in A : f(x) > t\}) \, dt. \tag*{□}
\]

Example 3.59. 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 3.60:

1. We show that

\[ I = \int_{-\infty}^{\infty} e^{-x^2} \, dx = \sqrt{\pi} \]

in two ways.

Note that \( e^{-x^2} > 0 \) for every \( x \in \mathbb{R} \) and

\[ I \leq \int_{-\infty}^{-1} -xe^{-x^2} \, dx + \int_{-1}^{1} e^{-x^2} \, dx + \int_{1}^{\infty} xe^{-x^2} \, dx < \infty. \]

Since \( x \mapsto e^{-x^2} \) is even,

\[ I = 2 \int_{0}^{\infty} e^{-x^2} \, dx. \]

Now

\[ I^2 = \left( \int_{-\infty}^{\infty} e^{-x^2} \, dx \right) \left( \int_{-\infty}^{\infty} e^{-y^2} \, dy \right) = 4 \int_{0}^{\infty} \left( \int_{0}^{\infty} e^{-(x^2+y^2)} \, dx \right) \, dy. \]

Substitution \( y = xs \) implies \( dy = x \, ds \). By Fubini’s theorem

\[ \frac{I^2}{4} = \int_{0}^{\infty} \left( \int_{0}^{\infty} e^{-(1+s^2)x^2} \, dx \right) \, ds = \frac{1}{2} \int_{0}^{\infty} \frac{1}{1 + s^2} \, ds = \frac{1}{2} \arctan(s) \bigg|_{0}^{\infty} = \frac{\pi}{4}. \]

Thus \( I = \sqrt{\pi} \).

2. \( \int_{\mathbb{R}^2} e^{-(x^2+y^2)} \, dx \, dy = \lim_{i \to \infty} \int_{B_i(0,1)} e^{-(x^2+y^2)} \, dx \, dy = \lim_{i \to \infty} \int_{B(0,1)} e^{-(x^2+y^2)} \, dx \, dy \)

\[ = \lim_{i \to \infty} \int_{0}^{\pi} \int_{0}^{2\pi} e^{-r^2} r \, dr \, d\theta = 2\pi \lim_{i \to \infty} \int_{0}^{i} e^{-r^2} r \, dr \]

\[ = \pi \lim_{i \to \infty} -e^{-r^2} \bigg|_{0}^{i} = -\pi \lim_{i \to \infty} (e^{-i^2} - 1) = \pi, \]

where all integrals are (possibly improper) 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

\[ \pi = \int_{\mathbb{R}^2} e^{-(x^2+y^2)} \, dx \, dy = \left( \int_{\mathbb{R}} e^{-x^2} \, dx \right) \left( \int_{\mathbb{R}} e^{-y^2} \, dy \right) \]

\[ = \int_{\mathbb{R}} e^{-x^2} \left( \int_{\mathbb{R}} e^{-y^2} \, dy \right) \, dx = \left( \int_{\mathbb{R}} e^{-x^2} \, dx \right)^2 \]

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 3.61:

(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\[
\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2 + y^2)^2} \, dy \, dx = \int_0^1 \frac{1}{x^2 + 1} \, dx = \arctan x \bigg|_0^1 = \frac{\pi}{4}.
\]
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
\[
\int_0^1 \int_0^1 \left| \frac{x^2 - y^2}{(x^2 + y^2)^2} \right| \, dy \, dx = \int_0^1 \int_0^1 \frac{y^2 - x^2}{(x^2 + y^2)^2} \, dy \, dx
\]
\[
= \int_0^1 \frac{1}{x} \, dx - \int_0^1 \frac{1}{x^2 + 1} \, dx = \infty.
\]
The integral in the brackets can be evaluated by integration by parts.

(2) Let \( f : [0,1] \times [0,1] \to \mathbb{R} \) be defined by
\[
f(x,y) = \begin{cases} 
2^{2i}, & 2^{-i} \leq x < 2^{-i+1}, 2^{-i} \leq y < 2^{-i+1}, \\
-2^{2i}, & 2^{-i-1} \leq x < 2^{-i}, 2^{-i} \leq y < 2^{-i+1}, \\
0, & \text{otherwise.}
\end{cases}
\]
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
\[
\int_{[0,1]} \left( \int_{[0,1]} f(x,y) \, dx \right) \, dy = 0,
\]
\[
\int_{[0,1]} \left( \int_{[0,1]} f(x,y) \, dy \right) \, dx = 1 \quad \text{and}
\]
\[
\int_{[0,1]} \left( \int_{[0,1]} |f(x,y)| \, dy \right) \, dx = \infty.
\]
(Exercise)

(3) Let \( 0 = \delta_1 < \delta_2 < \cdots < 1 \) and \( \delta_i \to 1 \) as \( i \to \infty \). Let \( g_i, i = 1,2,\ldots \) be continuous functions such that
\[
\text{supp} \, g_i \subset (\delta_i,\delta_{i+1}) \quad \text{and} \quad \int_0^1 g_i(t) \, dt = 1
\]
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
\[
\int_0^1 \int_0^1 |f(x,y)| \, dy \, dx = \infty.
\]
(4) Let
\[
Q = \{(x,y) \in \mathbb{R}^2 : x,y \geq 0\},
\]
\[
R = \{(x,y) \in Q : x-1 \leq y \leq x\},
\]
\[
S = \{(x,y) \in Q : x-2 \leq y \leq x-1\}.
\]
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
\[
g(x) = \int_{-\infty}^{\infty} f(x,y) \, dy = \begin{cases} 
-x, & 0 \leq x \leq 1, \\
x - 2, & 1 \leq x \leq 2, \\
0, & x \geq 2.
\end{cases}
\]
Then \( \int_{-\infty}^{\infty} g(x) \, dx = -1 \). Similarly
\[
h(y) = \int_{-\infty}^{\infty} f(x,y) \, dx = 0
\]
for every \( y \), but
\[
\int_{-\infty}^{\infty} h(y) \, dy = 0 \neq -1 = \int_{-\infty}^{\infty} g(x) \, dx.
\]


