MEASURE THEORY

JOHN E. HUTCHINSON
DEPARTMENT OF MATHEMATICS
SCHOOL OF MATHEMATICAL SCIENCES, A.N.U.

Introduction

- In the following we survey the main results in the theory of measure and integration. The main references I have used are [EG], [S] and [R], in that order.
- Proofs are usually only sketched, but I have attempted to provide a reasonable amount of motivation of both proofs and results.
- *We will often consider general measures μ on an arbitrary set X. But you should first think of the most important case — Lebesgue measure in $\mathbb{R}^n$. To fix ideas, take $n = 2$.*
- There are a considerable number of footnotes. I have done this so as not to distract from the main ideas. I suggest you avoid the footnotes on first reading and when reviewing the material

Contents

1 Lebesgue and Other Measures

1.1 Motivation

1. *Lebesgue measure*[1] is a way of assigning to arbitrary subsets of $\mathbb{R}^n$ a number which corresponds to the "size" of the set. Think of an infinite mass uniformly distributed over $\mathbb{R}^n$ such that the mass in any unit n-cube is one; the Lebesgue measure $\mathcal{L}(A)$ of a set A is the "amount of matter in A". In particular, $0 \le \mathcal{L}(A) \le \infty$.

   Lebesgue measure is the most important example of a measure; you should usually think of this case in the general theory which follows.

2. *Radon measures* form a very important class of measures. Lebesgue measure is a Radon measure. A Radon measure μ corresponds to a mass distribution in $\mathbb{R}^n$, where the amount of matter in any bounded set is finite. The measure $\mu(A)$ of $A \subset \mathbb{R}^n$ is again the "amount of matter in A". For example, μ may correspond to α units of mass concentrated at a point $P \in \mathbb{R}^n$. Then $\mu(A) = \alpha$ if $P \in A$ and $\mu(A) = 0$ otherwise. If μ corresponds to β units of mass uniformly distributed along a curve C of length one, then $\mu(A) = \beta \times$ (length of C in A).

3. *Borel regular measures* are the most general measures one usually considers in $\mathbb{R}^n$. Examples include the Radon measures. Other examples are k-dimensional *Hausdorff measure* $\mathcal{H}^k$, where $0 \le k \le n$. $\mathcal{H}^0$ is "counting measure" and gives the cardinality of a set; $\mathcal{H}^1$ is "length"; $\mathcal{H}^2$ is "area", ..., $\mathcal{H}^n$ is the same as Lebesgue measure in $\mathbb{R}^n$.[2] $\mathcal{H}^k$ is always Borel regular, but is not a Radon measure if $k < n$.[3]

1.2 Lebesgue Measure

1.2.1 Introduction

If $R \subset \mathbb{R}^n$ is a "rectangle", i.e.

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n],$$

then we define

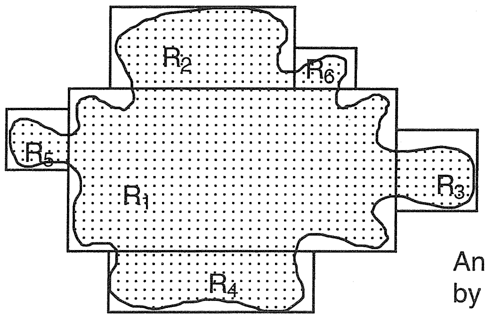
$$m(R) = (b_1 - a_1) \cdot \ldots \cdot (b_n - a_n).$$

For general $A \subset \mathbb{R}^n$ the definition of $\mathcal{L}(A)$, the Lebesgue measure of A, is motivated by the idea of covering A as "efficiently" as possible by rectangles.

[1]What we call a *measure* is often called an *outer measure*.

[2]One can also define Hausdorff measure $\mathcal{H}^k$ for non-integer k. This is useful for analysing "Cantor-type" sets.

[3]Why do we expect this?



An "efficient"cover by rectangles

Definition If $A \subset I\!R^n$ then the *Lebesgue measure* of A is defined by

$$\mathcal{L}^n(A) = \mathcal{L}(A) = \inf_{A \subset \bigcup_{i=1}^{\infty} R_i} \sum m(R_i),$$

where the R_i are rectangles.[4]

One can show that if R is a rectangle, then

$$\mathcal{L}(R) = m(R).$$

Note that "$\leq$" is immediate; "$\geq$" is basically a combinatorial argument.

1.2.2 Elementary Properties

The following properties of $\mathcal{L}$ are used to define the notion of a general measure. See Section 1.3.1.

1. $\mathcal{L}(\emptyset) = 0$,
2. $A \subset \bigcup_{i=1}^{\infty} A_i \Rightarrow \mathcal{L}(A) \leq \sum_{i=1}^{\infty} \mathcal{L}(A_i)$.

PROOF: *Exercise.* ∎

Exercises

1. Any singleton, and hence any countable set, has (Lebesgue) measure zero.
2. Any line segment in $I\!R^2$ has measure zero.

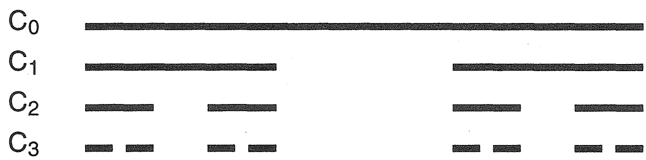
[4](a) One can consider "closed" rectangles, or half-open rectangles, etc., without changing the value of $\mathcal{L}(A)$. *Why?*
(b) Finite covers are allowed; just take $R_i = \emptyset$ for $i \geq N$, say.

Countable Additivity It seems reasonable to expect from our intuitive idea of Lebesgue measure that if $A = \bigcup_{i=1}^{\infty} A_i$ and the A_i are *mutually disjoint*, then $\mathcal{L}(A) = \sum_{i=1}^{\infty} \mathcal{L}(A_i)$. Unfortunately this is *not* true, but it *is* true for the so-called *(Lebesgue) measurable* sets. Essentially any set we come across in Analysis is (Lebesgue) measurable, as we will discuss later. See Sections 1.3.2 and 1.4.3.

1.2.3 Sets of Measure Zero

A set is *null* if it has measure zero.

We noted in Section 1.2.2 that any countable set is null. The *Cantor set* C is an example of an *uncountable* null set.



Approximations to
the Cantor set

One constructs C as follows:

$$\begin{aligned} C_0 &= [0,1] \\ C_1 &= [0,1/3] \cup [2/3,1] \\ C_2 &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\ &\vdots \\ C &= \bigcap_{i=0}^{\infty} C_i. \end{aligned}$$

Thus C_i consists of 2^i closed intervals of length 3^{-i}; and C_{i+1} is obtained from C_i by removing the middle (open) third of each interval in C_i.

By the definition of $\mathcal{L}$,

$$\mathcal{L}(C) \le 2^i \times 3^{-i}$$

for all i. Hence

$$\mathcal{L}(C) = 0.$$

The Cantor set C is uncountable[5] since there is a one-one correspondence[6] between elements of C and those reals in $[0,1]$ which have a ternary expansion containing only the numerals 0 and 2. By replacing 2 by 1 and considering binary expansions, we obtain a map from the set of such reals onto $[0,1]$.[7]

[5] C does *not* consist only of endpoints of the intervals C_i.

[6] What is it?

[7] The map is not one-one, because of the non-uniqueness of the binary expansion of a real number.

1.3 General Measures

It is now convenient to generalise our previous considerations. *In the following think of the case* $X = I\!R^2$.

1.3.1 Introduction

The following definition is motivated by the two properties for Lebesgue measure noted in Section 1.2.2.

Definition A *measure* μ on a set X is a function which assigns to every $A \subset X$ a number $\mu(A) \in [0, \infty]$ such that[8]

$$\mu(\emptyset) = 0,$$
$$A \subset \cup_{i=1}^{\infty} A_i \Rightarrow \mu(A) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

Thus Lebesgue measure is a measure on $I\!R^n$ from Section 1.2.2.

Definition If $A \subset X$ then the measure μ *restricted to* A is the measure $\mu \lfloor A$ on X defined by

$$\mu \lfloor A\,(B) = \mu(A \cap B).$$

It is straightforward to check that $\mu \lfloor A$ is a measure on X (*not* on A). If we think of μ as corresponding to a mass distribution, then $\mu \lfloor A$ is the mass distribution obtained by removing any matter outside A.[9]

1.3.2 Measurable Sets

If μ is a measure then we define the class of *μ-measurable sets* in such a way that μ is "countably additive" on this class. The following definition is due to Caratheodory. It is not immediately clear where it comes from; it's virtue is that it "works".[10]

Definition A set A is (μ-)*measurable* if for any set B,

$$\mu(B) = \mu(B \cap A) + \mu(B \cap A^c).^{11} \tag{1}$$

---

[8](a) It follows that if $A \subset A_1 \cup \cdots \cup A_N$ then $\mu(A) \leq \mu(A_1) + \cdots + \mu(A_N)$. Just take $A_i = \emptyset$ if $i > N$ in the Definition.
(b) In particular, if $A \subset B$ then $\mu(A) \leq \mu(B)$.

[9]One could also define a measure $\mu|_A$ on A, rather than X, by $\mu|_A(B) = \mu(B)$ for all $B \subset A$.

[10]The idea is that A *fails* to be measurable if it is so badly intertwined with its complement that (1) fails to be true for some B.

[11](a) $B^c = X \setminus B$ is the *complement* of B.
(b) Thus A is measurable iff it splits every set additively.

Note that "$\leq$" is true from the first Definition in Section 1.3.1.

It is immediate that

1. $\emptyset$ and X are measurable, and
2. any set of measure zero is measurable.

For a general measure, only X and $\emptyset$ need be measurable.[12] We will see later that any set we are likely to encounter in Analysis is Lebesgue measurable, c.f. Section 1.4.3.

The class of μ-measurable sets forms a *σ-algebra*, i.e. is closed under complements, countable intersections and countable unions. Moreover, μ has precisely the properties on this class that we might expect from our intuition about a "measure".

Proposition *Let A and $A_1, A_2, \ldots$ be measurable. Then*

1. *A^c, $\bigcup_{i=1}^{\infty} A_i$ and $\bigcap_{i=1}^{\infty} A_i$ are measurable;*
2. *if A_i are disjoint then*
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i);$$
3. *if $A_1 \subset A_2 \subset \cdots$ then*
$$\mu(A_i) \uparrow \mu\left(\bigcup_{i=1}^{\infty} A_i\right);$$
4. *if $A_1 \supset A_2 \supset \cdots$ and $\mu(A_1) < \infty$[13] then*
$$\mu(A_i) \downarrow \mu\left(\bigcap_{i=1}^{\infty} A_i\right).$$

The proof is elementary, but not trivial.

Exercise If Ir is the set of irrationals then Ir is Lebesgue measurable and $\mathcal{L}(Ir) = \infty$.

(c) If A is not measurable, then (1) will in fact fail with some "nice" set B; e.g. some rectangle B in the case of Lebesgue measure.

[12] Consider the measure $\mu(\emptyset) = 0$ and $\mu(A) = 1$ otherwise.

[13] This condition is necessary; let $A_i = [i, \infty)$.

Almost Everwhere If μ is a measure on X we say that a property P holds *almost everywhere* if it holds except on a set of (μ-)measure zero. We write "P holds μ a.e.", or "P holds a.e." if μ is understood from context.[14]

Sets of measure zero (null sets) can usually be ignored in measure theory.

1.4 Measures on $\mathbb{R}^n$

To fix ideas, think of $\mathcal{L}$ on $\mathbb{R}^2$.

1.4.1 Borel Regular Measures

Definition A measure μ on $\mathbb{R}^n$ is *Borel regular* if

1. every Borel[15] set is μ-measurable,
2. for *every* set A there is a Borel set $B \supset A$ such that $\mu(B) = \mu(A)$.[16]

Proposition *Lebesgue measure is a Borel regular measure.*

PROOF: One uses Caratheodory's criterion, which says it is sufficient to prove $\mathcal{L}(A \cup B) = \mathcal{L}(A) + \mathcal{L}(B)$ whenever the distance between A and B is > 0. This equality is fairly straightforward to prove. ∎

The following is a characterisation of the Lebesgue measurable sets.

Proposition *A set A is Lebesgue measurable $\iff$ $A = B \setminus N$ for some Borel set B and null set N $\iff$ $A = B \cup N$ for some Borel set B and null set N.*

PROOF: *Exercise* (use the previous Proposition). ∎

1.4.2 Radon Measures

Definition A *Radon measure* is a Borel regular measure such that every compact set has finite measure.

---

[14] In probability theory, one says P holds *almost surely* and writes "P holds a.s.".

[15] Recall that the class of *Borel sets* (c.f. Section 1.3.2) is the smallest σ-algebra containing the open sets. In particular, closed sets are Borel.

[16] B is called a *Borel cover* of A. Thus the set A can be "approximated from above" by a Borel set. If the second condition is also true with "$\supset$" replaced by "$\subset$", then it follows A is measurable. *Exercise.*

See Section 1.1 for examples and non-examples. In particular, Lebesgue measure is a Radon measure. If $a \in I\!R^n$ then the Radon measure corresponding to a unit mass at a is called the *Dirac measure concentrated at a* and is denoted by δ_a. Thus

$$\delta_a(E) = 1 \text{ if } a \in E, \ \ \delta_a(E) = 0 \text{ if } a \notin E.$$

Remark If μ is a Borel regular measure on $I\!R^n$, A is μ-measurable and $\mu(A) < \infty$, then $\mu \lfloor A$ is a Radon measure. This is straightforward to check.

An example is $\mathcal{H}^1$ measure on $I\!R^2$ restricted to a curve of finite length.

1.4.3 Lebesgue Measurable v. Non-Measurable Sets

The existence of Lebesgue non-measurable sets is proved using the uncountable axiom of choice. In fact it was proved by Solovay that one actually *requires* the *uncountable* axiom of choice in the construction. More precisely, it is consistent with the usual axioms of set theory, including the countable axiom of choice, that all subsets of $I\!R^n$ are Lebesgue measurable.

Moreover, we have seen that the Borel sets are Lebesgue measurable, and any sets constructed from the Borel sets by any finite or countable set theoretic operation are also Lebesgue measurable. If a set is constructed from measurable sets by some sort of limiting operation which is not countable, one can often use the density of the rationals to give another "countable" construction of the set, and thus deduce its measurability.

The moral of all this is that you need not worry about non-measurable sets, at least when working with Lebesgue measure, or more generally with Borel regular measures.[17]

1.5 Approximation Results

Theorem *Suppose μ is a Radon measure. Then*

1. *for each $A \subset I\!R^n$*

$$\mu(A) = \inf\{\mu(U) : A \subset U, \ U \textit{ open}\};$$

2. *for each μ-measurable $A \subset I\!R^n$*

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

[17] In probability theory, and particularly in the theory of stochastic processes, it is natural to identify the measurable sets with the class of events which are observable in some sense. In this case, non-measurable sets can be quite important.

Thus A can be approximated from the outside by open sets and (if it is measurable) from the inside by compact sets.[18]

PROOF: The proof proceeds in steps.

1. Let

$$\begin{aligned} \mathcal{F} = \{A : A &\text{ is measurable, and for each } \epsilon > 0 \text{ there exists} \\ &\text{a closed set } C \subset A \text{ such that } \mu(A \setminus C) < \epsilon\} \\ \mathcal{G} = \{A &\in \mathcal{F} : A^c \in \mathcal{F}\}. \end{aligned}$$

   Then one checks that

   (a) $\mathcal{F}$ contains all closed sets
   (b) $\mathcal{F}$ is closed under countable intersections
   (c) $\mathcal{F}$ is closed under countable unions
   (d) $\mathcal{F}$ contains all open sets
   (e) $\mathcal{G}$ is closed under complements and countable unions
   (f) $\mathcal{G}$ contains the open sets and hence all Borel sets.

   In particular, each Borel set B contains a closed set C such that $\mu(B \setminus C) < \epsilon$

2. For each Borel B there is an open $U \supset B$ such that $\mu(U \setminus B) < \epsilon$ (this follows by applying the previous result to $B_N(0) \setminus B$ for large N)[19].

3. The first claim of the Theorem is next established for Borel sets, and hence for arbitrary sets using Borel regularity.

4. The final result follows from the previous step, essentially by taking complements, but there are a few technical points.

■

[18](a) One cannot approximate from the *inside* by open sets; let A be the set of irrationals. The only open subset is $\emptyset$. Similarly, one cannot approximate from the *outside* by compact sets, even if A is bounded; let A be the set of rationals in $[0, 1]$.
(b) Measurability is needed in *2* of the Theorem; in fact equality holds iff A is measurable.
(c) It does *not* follow that for each $\epsilon > 0$ there is an open set $U \supset A$ such that $\mu(U \setminus A) < \epsilon$. But this *is* true if A is measurable.
(d) It also follows that if A is measurable then there is a *closed* $C \subset A$ such that $\mu(A \setminus C) < \epsilon$. This is not true for compact C; take $A = \mathbb{R}^n$.

[19] $B_N(0) := \{x \in \mathbb{R}^n : |x| < N\}$.

2 Measurable Functions and Integration

In this Section, μ is a measure on the set X.

Think of the case $X = I\!R^n$, $\mu = \mathcal{L}$ *(and* $Y = I\!R$*).*

2.1 Measurable Functions

2.1.1 Introduction

Essentially any function $f : I\!R^n \to I\!R$ which arises in Analysis will be measurable with respect to a given Borel regular measure. Moreover, the class of measurable functions is closed under finite and countable operations.

Definition Let μ be a measure on a set X, Y be a topological space, and $f : X \to Y$. Then f is *μ-measurable* if for any open set $U \subset Y$, $f^{-1}[U]$ is μ-measurable.

It follows (*Exercise*) that $f^{-1}[B]$ is μ-measurable for any closed set B, and more generally for any Borel set B.

In case $Y = I\!R$, f is μ-measurable iff (*Exercise*)

$$f^{-1}(-\infty, a] := \{x : f(x) \leq a\} \text{ is } \mu\text{-measurable for all real numbers } a\ . \quad (2)$$

Similarly, one can instead consider intervals of the form $(-\infty, a)$, or (a, b), or $(a, b]$, etc.

Remark It is often convenient to consider functions $f : X \to [-\infty, +\infty]$. Such a function is said to be *measurable* if (2) holds and if both of the sets $\{x : f(x) = -\infty\}$ and $\{x : f(x) = +\infty\}$ are measurable.

2.1.2 Elementary Properties

Proposition *Suppose* $f, g, f_1, f_2, \ldots : X \to [-\infty, +\infty]$ *are real-valued μ-measurable functions, and α is a real number. Then the following are μ-measurable:*[20] [21]

$$\alpha f,\ f+g,\ fg,\ f/g,\ |f|,\ \min\{f,g\},\ \max\{f,g\},$$
$$\inf_i f_i,\ \sup_i f_i,\ \liminf_{i\to\infty} f_i,\ \limsup_{i\to\infty} f_i,\ \lim_{i\to\infty} f_i.$$

[20] Of course, the relevant expressions must be defined. Thus we need to either define $\infty - \infty$ and $0/0$, or to consider only those f and g where this does not happen. We also assume that the final limit exists.

[21] If $(a_i)_{i=1}^{\infty}$ is a sequence of real numbers then

$$\liminf_{i\to\infty} a_i := \lim_{k\to\infty} \inf_{i\geq k} a_i .$$

The proof is routine, using (2).

2.1.3 Littlewood's Three Principles

For Lebesgue measure (or more generally, Borel regular measures) one has[22]

- *Every (measurable) set is nearly open;*
- *Every (measurable) function is nearly continuous;*
- *Every pointwise convergent sequence of (measurable) functions is nearly uniformly convergent.*

We have seen a version of the first principle in the Theorem in Section 1.5. For the second and third principle see Lusin's Theorem in Section 2.1.4 and Egoroff's Theorem in Section 2.1.5 respectively.

It is frequently the case that if a result is true for open sets, continuous functions, or uniform convergence respectively, then some version of the corresponding principle enables one to establish the result in general.

2.1.4 Lusin's Theorem

Theorem *Suppose μ is a Radon measure on $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is μ-measurable. Then for each $\epsilon > 0$ there is a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ such that $f = g$ except on a set of μ-measure less than ϵ.*

PROOF: First assume $\mu(\mathbb{R}^n) < \infty$.

1. For each integer i find a partition of $\mathbb{R}^n$ into measurable sets such that f varies by at most $1/i$ on each member of the partition.
2. Approximate members of the partition by disjoint compact *subsets* to within a total μ-measure error $\epsilon/2^i$, using the Theorem in Section 1.5.
3. On these compact sets approximate f by a function f_i which is constant on each compact set.
4. The f_i converge uniformly on their common domain D to a function g^* which is continuous on D and agrees with f on D.

Equivalently, $a = \liminf_{i\to\infty} a_i$ iff a is the least extended real number, possibly $-\infty$, for which there is a subsequence $a_{i'} \to a$. Similarly for lim sup. Also

$$\left(\liminf_{i\to\infty} f_i\right)(x) := \liminf_{i\to\infty} f_i(x).$$

[22] c.f. Littlewood, *Lectures on the Theory of Functions*, Oxford, 1944, p. 26.

5. By Tietze's Extension Theorem, extend g^* to a continuous function g defined on all of $\mathbb{R}^n$.

The case $\mu(\mathbb{R}^n) = \infty$ follows by considering $\mu \lfloor B_R$ and letting $R \to \infty$. ∎

2.1.5 Egoroff's Theorem

Theorem *Suppose μ is a measure on $\mathbb{R}^n$ and $f_i : \mathbb{R}^n \to \mathbb{R}$ are μ-measurable functions. Suppose $A \subset \mathbb{R}^n$ is μ-measurable, $\mu(A) < \infty$ and $f_i \to f$ a.e. on A.*[23] *Then for each $\epsilon > 0$ there is a μ-measurable set $B \subset A$ such that $\mu(A \setminus B) < \epsilon$ and $f_i \to f$ uniformly on B.*

PROOF: This is a general result holding for arbitrary measures; and the proof is fairly straightforward

1. First show that for each $\delta > 0$ and $\epsilon > 0$ there exists an integer N and a set $A_{\delta,\epsilon} \subset A$ with $\mu(A \setminus A_{\delta,\epsilon}) < \epsilon$ such that $|f(x) - f_i(x)| < \delta$ if $x \in A_{\delta,\epsilon}$ and $i \geq N$.
2. Let $B = \bigcup_{i=1}^{\infty} A_{1/i,\,\epsilon/2^i}$.

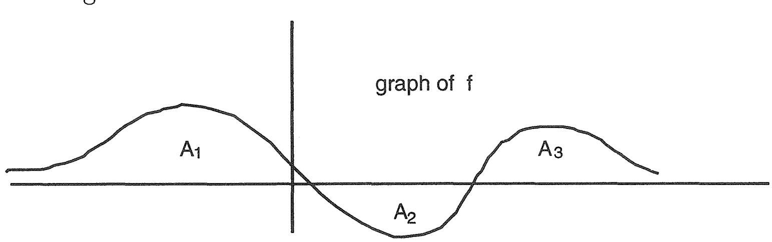
∎

The condition $\mu(A) < \infty$ is necessary. Let $f_i(x) = x/i$ for $x \in \mathbb{R}$, $A = \mathbb{R}$ and $\mu = \mathcal{L}$.

2.2 Integration

2.2.1 Introduction

The idea is that if $f : \mathbb{R}^n \to [-\infty, \infty]$ then $\int f \, d\mathcal{L}^n$ is the signed $\mathcal{L}^{n+1}$ measure of the volume under the graph, taking the region below the $\mathbb{R}^n$ plane with a negative sign.



$\int f = L^2(A_1) + L^2(A_3) - L^2(A_2)$

[23] That is, $f_i(x) \to f(x)$ for $x \in A \setminus N$, where $\mu(N) = 0$.

However, this approach requires $\mathcal{L}^{n+1}$ measure in order to define integration with respect to $\mathcal{L}^n$ (and the "product measure" $\mu \times \mathcal{L}^1$ to define integration with respect to μ).

Instead, we first define the integral of positive simple functions, then of arbitrary positive (measurable) functions, then of arbitrary (measurable) functions.

Definition Suppose μ is a measure on the set X and $f : X \to [-\infty, \infty]$ is measurable.

1. If $f = \sum_{i=1}^{N} \alpha_i \chi_{E_i}$ [24] where $\alpha_i \geq 0$ and the E_i are disjoint measurable sets[25], then f is said to be a *positive simple function* and
$$\int f\, d\mu = \alpha_1 \mu(E_1) + \cdots + \alpha_N \mu(E_N).$$
2. If $f \geq 0$ then
$$\int f\, d\mu = \sup\left\{ \int u\, d\mu : u \leq f,\ u \text{ is positive simple} \right\}.$$
3. For arbitrary measurable f,
$$\int f\, d\mu = \int f^+\, d\mu - \int f^-\, d\mu,^{26}$$
provided it is not the case that both terms on the right are $+\infty$.

The function f is *integrable* if $\int f\, d\mu$ is well-defined.[27] The function f is *summable* if $\int f\, d\mu$ exists and is *finite*.[28]

If E is μ-measurable, then we define *the integral of f over E* by
$$\int_E f\, d\mu = \int f \chi_E\, d\mu.$$

The function f is *locally summable* if $\int_E |f|\, d\mu < \infty$ for all compact E.[29] [30]

[24] χ_E is the *characteristic function* of E and equals one on E and zero otherwise.

[25] We could drop the positivity and disjointedness conditions, but these are the only cases we need for the rest of the definition.

[26] Where the positive and negative parts of f are defined by
$$f^+ := \max\{f, 0\},\ f^- := -\min\{f, 0\}.$$

[27] Thus the *only* way a measurable function can fail to be integrable is if both the positive and negative parts have integral equal to ∞.

[28] Equivalently, $\int |f|\, d\mu < \infty$. Many texts use "integrable" for what we call "summable".

[29] Thus the constant function one is locally summable but not summable with respect to Lebesgue measure. The function $1/x$ defined on $\mathbb{R}$ is not even locally summable.

[30] See also Section 4.2.

Remarks

1. One can show that the definition is consistent, in that a positive simple function has the same integral by any of the three definitions and a positive function has the same value by either of the last two definitions.

2. If $f \geq 0$ takes the value ∞ on a set of positive measure, then $\int f = \infty$.

3. If $f = g$ a.e. then f is integrable iff g is integrable, and in this case $\int f = \int g$. In particular, if $f = 0$ except on a set of measure 0, then $\int f = 0$.

Riemann Integration It is not difficult to show that if $f : [a, b] \to I\!R$ is Riemann integrable, then the Lebesgue integral $\int_{[a,b]} f$ exists and has the same value.

It is interesting to note that the Riemann integral of f is defined by partitioning the *domain* of f into intervals and passing to a limit. On the other hand, the Lebesgue integral is essentially defined by dividing the *range* of f into intervals, considering the corresponding pre-image sets in $I\!R^n$, and passing to an appropriate limit.

2.2.2 Elementary Properties

One has

$$
\begin{aligned}
\int \alpha f &= \alpha \int f \qquad (\alpha \in I\!R) \\
\int f + g &= \int f + \int g \\
f \leq g &\Rightarrow \int f \leq \int g \\
\int_E f + \int_F f &= \int_{E \cup F} f \qquad (E \text{ and } F \text{ disjoint}).
\end{aligned}
$$

More precisely, it is assumed that f and g are integrable, and in the second case that the sum on the right is defined. Also E and F are measurable.

In future, all sets and functions are assumed measurable with respect to the relevant measure. All integrals are assumed to be well-defined.

2.3 Limit Theorems

2.3.1 Discussion

Much of the importance of Lebesgue integration is a consequence of the fact that under "reasonable" conditions,

$$
f_i \to f \text{ (a.e.)} \Rightarrow \int f_i \to \int f. \tag{3}
$$

Note that if such a result is true for "everywhere" convergence, then it is also true for "a.e." convergence.[31]

The following three examples show the type of behaviour we need to avoid and show the necessity of the various hypotheses in the subsequent three theorems. Rougly speaking, if we can eliminate the following problems then (3) will hold.

Example 1 Let

$$\begin{aligned} f_n(x) &= \begin{cases} n & 0 < x \leq 1/n \\ 0 & \text{otherwise} \end{cases} \\ f(x) &= \quad 0 \qquad\qquad \text{all } x \end{aligned}$$

Then $f_n \to f$ everywhere, but $\int f_n \, d\mathcal{L} = 1$ and $\int f \, d\mathcal{L} = 0$.

Example 2 Let

$$\begin{aligned} f_n(x) &= \begin{cases} 1 & n \leq x \leq n+1 \\ 0 & \text{otherwise} \end{cases} \\ f(x) &= \quad 0 \qquad\qquad \text{all } x \end{aligned}$$

Then $f_n \to f$ everywhere, but $\int f_n \, d\mathcal{L} = 1$ and $\int f \, d\mathcal{L} = 0$.

Example 3 Let

$$\begin{aligned} f_n(x) &= -1/n \qquad \text{all } x \\ f(x) &= \quad 0 \qquad \text{all } x \end{aligned}$$

Then $f_n \uparrow f$ everywhere, but $\int f_n \, d\mathcal{L} = -\infty$ and $\int f \, d\mathcal{L} = 0$.

2.3.2 Fatou's Lemma

Theorem *Suppose $f_i : X \to [-\infty, \infty]$ are μ-measurable for $i = 1, 2, \ldots$ and $f_i \geq g$ for some μ-summable function g. Then*

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

Remark The most commonly used form is

$$0 \leq f_i \to f \quad \Rightarrow \quad \int f \, d\mu \leq \liminf_{i\to\infty} \int f_i \, d\mu.$$

[31] By redefining the f_i on a set of measure zero, one obtains covergence everywhere, but does not change the value of any of the integrals involved.

Example 3 in Section 2.3.1 shows the need for the bound from below in the hypotheses; Example 1 shows that one can only expect "$\leq$" in the conclusion.

PROOF: By subtracting g from the f_i we assume $f_i \geq 0$.

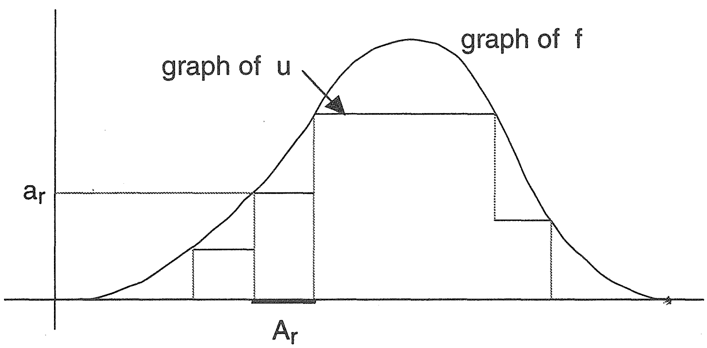
Let

$$f = \liminf_{i\to\infty} f_i$$

(think of the case where $\lim_{i\to\infty} f_i$ exists).

Fix $\epsilon > 0$. Choose

$$u = \sum_{r=1}^{m} a_r \chi_{A_r} \geq 0, \ A_r \text{ disjoint}, \ \int u \geq \int f - \epsilon.$$



Fix $0 < t < 1$ (think of t as near 1). Let

$$B_{r,k} = \{x \in A_r : f_j(x) \geq ta_r \ \forall j \geq k\}.$$

Then

$$B_{r,k} \uparrow A_r \text{ and so } \mu(B_{r,k}) \uparrow \mu(A_r), \text{ as } k \to \infty.$$

Hence for each k,

$$\begin{aligned} \int f_k &\geq \sum_r \int_{A_r} f_k \\ &\geq \sum_r \int_{B_{r,k}} f_k \\ &\geq \sum_r \int_{B_{r,k}} ta_r \\ &= t\sum_r a_r \mu(B_{r,k}). \end{aligned}$$

Letting $k \to \infty$,

$$\liminf_{k\to\infty} \int f_k \geq t \sum_r a_r \mu(A_r) = t \int u.$$

Since t can be taken arbitrarily close to 1,

$$\liminf_{k\to\infty} \int f_k \geq \int u.$$

Since ϵ can be taken arbitrarily close to 0,

$$\liminf_{k\to\infty} \int f_k \geq \int f.$$

∎

2.3.3 Monotone Convergence Theorem

Theorem *Suppose $f_i : X \to [-\infty, \infty]$ are μ-measurable for $i = 1, 2, \ldots$ and $f_i \uparrow f$ a.e. Then*

$$\int f_i \, d\mu \uparrow \int f \, d\mu.$$

PROOF: Clearly $\int f_i \uparrow$, $\lim \int f_i$ exists (possibly $+\infty$), and $\lim \int f_i \leq \int f$. The reverse inequality follows from Fatou's lemma. ∎

Simple Application

$$\int_{[0,1]} x^\alpha dx = \begin{cases} \frac{1}{\alpha+1} & \alpha > -1 \\ \infty & \alpha \leq -1 \end{cases}$$

PROOF: Note that x^α is unbounded and so not Riemann Integrable. But if

$$f_n(x) = \begin{cases} x^\alpha & \frac{1}{n} \leq x \leq 1 \\ 0 & 0 \leq x < \frac{1}{n} \end{cases},$$

then

$$\int_{[0,1]} f_n = \int_{[1/n,1]} f_n = \begin{cases} \frac{1}{\alpha+1}\left(1 - \left(\frac{1}{n}\right)^{\alpha+1}\right) & \alpha \neq 1 \\ -\log\left(\frac{1}{n}\right) & \alpha = -1 \end{cases}$$

(using standard rules for Riemann integration).

Since $f_n \uparrow x^\alpha$, the result follows from the monotone convergence theorem. ∎

Rule of Thumb Integrals like the preceding cause no problems. Just use your commonsense in evaluating them.

2.3.4 Dominated Convergence Theorem

Theorem *Suppose* $f_i : X \to [-\infty, \infty]$ *are* μ*-measurable for* $i = 1, 2, \ldots$ *and* $|f_i| \le g$ *for some* μ*-summable function* g*. Suppose* $f_i \to f$ *a.e. Then*

$$\int f_i \, d\mu \to \int f \, d\mu.$$

PROOF: From Fatou's lemma,

$$\int f \le \liminf \int f_i$$

and

$$\int -f \le \liminf \int -f_i.$$

The last inequality implies

$$\int f \ge \limsup \int f_i.$$

This gives the result. ∎

3 Some Important Theorems

We consider some of the major results of measure theory. They will be used both here and in the other lecture series.

3.1 Product Measures

Think of the case $\mu = \nu = \mathcal{L}^1$, $X = Y = I\!R$. In this case one can show $\mu \times \nu = \mathcal{L}^2$ on $I\!R^2$.

The diagram in Section 1.2.1 is relevant to the following definition. Note that unlike in Section 1.2.1, even if $X = Y = I\!R$, the sets A_i and B_i need not be intervals.

Definition Let μ be a measure on X and ν be a measure on Y. Then the *product measure* $\mu \times \nu$ on $X \times Y$ is defined by

$$(\mu\times\nu)(S) = \inf\left\{\sum_{i=1}^{\infty}\mu(A_i)\nu(B_i) : S \subset \bigcup_{i=1}^{\infty}(A_i \times B_i),\ A_i \text{ and } B_i \text{ measurable}\right\}.$$

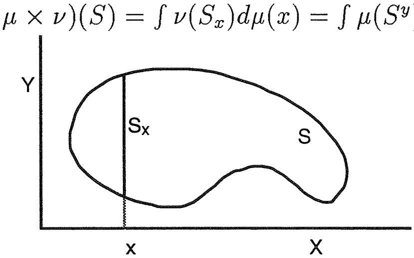
The following Theorem is not surprising. See [EG] for the (fairly straightforward, but long) proof.

Fubini's Theorem *Let μ be a σ-finite measure*[32] *on X and ν be a σ-finite measure on Y.*

1. *If $A \subset X$ and $B \subset Y$ are measurable, then $A \times B$ is measurable and $(\mu \times \nu)(A \times B) = \mu(A) \times \nu(B)$.*

2. *If μ and ν are Radon measures on $I\!R^m$ and $I\!R^n$ respectively, then $\mu \times \nu$ is a Radon measure on $I\!R^{m+n} \cong I\!R^m \times I\!R^n$.*

3. *If $S \subset X \times Y$ is $\mu \times \nu$ measurable then*
   $S_x := \{y : (x, y) \in S\}$ is measurable for a.e. x,
   $S^y := \{x : (x, y) \in S\}$ is measurable for a.e. y,
   $(\mu \times \nu)(S) = \int \nu(S_x)d\mu(x) = \int \mu(S^y)d\nu(y)$.



[32] A measure μ on X is *σ-finite* if there exist sets $(E_i)_{i=1}^{\infty}$ such that $\mu(E_i) < \infty$ and $X = \bigcup_{i=1}^{\infty} E_i$. In particular, *any Radon measure on $I\!R^n$ is σ-finite.*

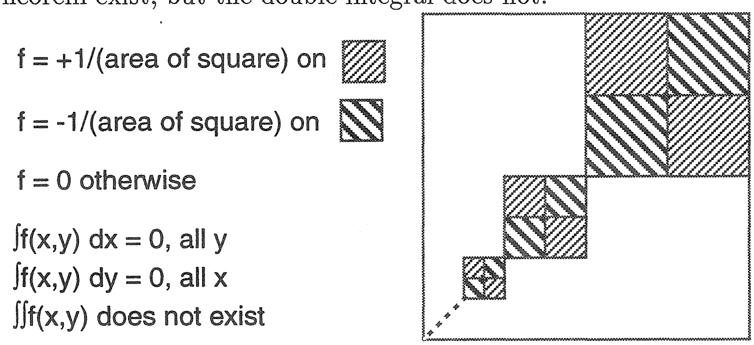
4. If f is $\mu \times \nu$-integrable[33] *then*
$\int f(x,y)\,d\nu(y)$ exists for a.e. x and is (measurable and) integrable
$\int f(x,y)\,d\mu(x)$ exists for a.e. y and is (measurable and) integrable,
$\int_{X \times Y} f(x,y) = \int (\int f(x,y)\,dy)\,dx = \int (\int f(x,y)\,dx)\,dy$.

Remark The main point to Fubini's Theorem is that one can evaluate a "double" integral by evaluating two "single" integrals.

The two main hypotheses to verify are (i) the *measurability* of f and (ii) the *existence* of $\int f\,d(\mu \times \nu)$. The first hypothesis is essentially always true in practice, as discusssed in Section 2.1.1.

The second hypothesis is true if $f \geq 0$ a.e. Alternatively, since Fubini's Theorem *always* applies to $|f|$ (being positive, and assuming as usual that f is measurable), one can often use Fubini to show that $\int |f|\,d(\mu \times \nu) < \infty$. But then $\int f\,d(\mu \times \nu)$ exists (and is finite), and hence Fubini can also be applied to f.

It is possible to find examples where both single integrals in *4* of Fubini's Theorem exist, but the double integral does not:



3.2 Change of Variable Formula

Definition Suppose $\phi\colon \mathbb{R}^n \to \mathbb{R}^n$ is C^1. The *Jacobian of ϕ at x* is defined by

$$\begin{aligned} J\phi(x) &= \text{absolute value of } \det[D\phi(x)] \\ &= \text{absolute value of } \det \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \cdots & \frac{\partial \phi_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial x_1} & \cdots & \frac{\partial \phi_n}{\partial x_n} \end{bmatrix} \end{aligned}$$

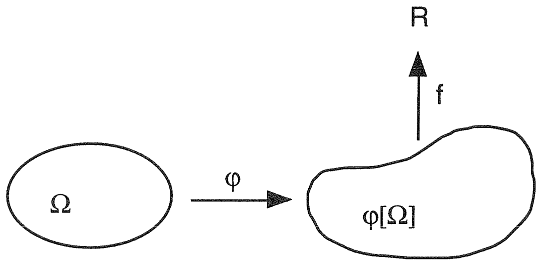
Geometrically, $J\phi(x)$ is the "volume expansion factor" for ϕ near x. This makes the following result geometrically plausible.

[33]Recall that this means the integral exists, possibly equal to $\pm\infty$.

Theorem *Suppose $\phi:\Omega \to \mathbb{R}^n$ is one-one and C^1 for some open $\Omega \subset \mathbb{R}^n$. Suppose $f:\phi[\Omega] \to \mathbb{R}$ and $f \in L^1(\phi[\Omega])$. Then*

$$\int_{\phi[\Omega]} f(y)\, dy = \int_\Omega f(\phi(x)) J\phi(x)\, dx.$$

In particular, the integral on the right exists and is finite.



PROOF: See [EG; Section 3.3.3]; the result is a consequence of the "Area Formula".[34] The proof is essentially done by first reducing to the case that f is a characteristic function and then splitting the domain into sets on which ϕ is "almost a polynomial of degree one". If ϕ is exactly a polynomial of degree one, the result is essentially linear algebra. ∎

The result is easy to remember, just formally replace y by $\phi(x)$, so in particular $dy = J(\phi(x))\, dx$.

3.3 Lebesgue Decomposition Theorem

Suppose μ and ν are Radon measures on $\mathbb{R}^n$.

Definition We say μ is *absolutely continuous* with respect to ν, and write

$$\mu \ll \nu,$$

if $\nu(E) = 0$ implies $\mu(E) = 0$ for all $E \subset \mathbb{R}^n$.

We say μ and ν are *mutually singular* and write

$$\mu \perp \nu,$$

if there exists a Borel set B such that

$$\mu(B) = \nu(B^c) = 0.$$

Note that two measures are mutually singular if they are "concentrated" on disjoint sets.

We will be most interested in these notions when $\nu = \mathcal{L}$.

[34]The result in [EG] is more general. The current result follows by replacing f and u in [EG] by ϕ and $f \circ \phi$ respectively.

Examples The "point mass" and "curve mass" examples in $\mathbb{R}^2$ from Section 1.1 are mutually singular to $\mathcal{L}^2$ and to each other.

If $f : \mathbb{R}^n \to \mathbb{R}$ is locally summable (see Section 2.2.1) and $f \geq 0$ then $f \lfloor \mathcal{L}$ is the Radon measure defined by

$$(f \lfloor \mathcal{L})(E) = \int_E f \, d\mathcal{L}. \tag{4}$$

Think of $f \lfloor \mathcal{L}$ as "Lebesgue measure weighted by the function f". Note that $f \lfloor \mathcal{L} \ll \mathcal{L}$, *why?*

The next theorem shows that any Radon measure on $\mathbb{R}^n$ can be decomposed into an absolutely continuous part and a singular part with respect to Lebesgue measure. The decomposition is essentially unique, the absolutely continuous part can be written as in (4), and the function f can be found "explicitly".

Theorem *Let μ be a Radon measure on $\mathbb{R}^n$. Then there exist unique Radon measures μ_{ac} and μ_S such that*

$$\mu = \mu_{ac} + \mu_S, \ \mu_{ac} \ll \mathcal{L}, \ \mu_S \perp \mathcal{L}.$$

Moreover,

$$\mu_{ac} = \frac{d\mu}{d\mathcal{L}} \lfloor \mathcal{L},$$

where

$$\frac{d\mu}{d\mathcal{L}}(x) = \lim_{r \to 0} \frac{\mu(B_r(x))}{\omega_n r^n}{}^{35} \quad \textit{for } \mathcal{L} \textit{ a.e. } x \tag{5}$$

is called the Radon-Nikodym derivative of μ with respect to $\mathcal{L}$, and is locally summable with respect to $\mathcal{L}$.

PROOF: The proof is via the Vitali Covering Theorem, which we do not have time to present. But see the lectures by Marty Ross. ∎

Generalisations There is an analogous result if $\mathcal{L}$ is replaced by an arbitrary Radon measure ν. The proof then requires the so-called Besicovitch Covering Theorem.

For more general measures on an arbitrary set, provided the measures are σ-finite,[36] an analogous theorem still holds. The major difference is that the existence of the Radon Nikodym derivative is obtained by a more abstract argument, and the concrete representation in (5) is no longer valid.

[35] By definition, $\omega_n = \mathcal{L}(B_1(0))$, and so $\omega_n r^n = \mathcal{L}(B_r(x))$ from the scaling and translation properties of Lebesgue measure.

[36] See footnote 32.

3.4 Lebesgue Points

Theorem *If $f: I\!\!R^n \to I\!\!R$ is locally summable then*

$$\lim_{r\to 0} \fint_{B_r(x)} f = f(x) \tag{6}$$

for a.e. x.

We say x is a *Lebesgue point* of f if (6) holds.

If $f = \chi_A$ is the characteristic function of the (measurable) set A, the theorem says that a.e. $x \in A$ is a *point of density one* and a.e. $x \notin A$ is a *point of density zero. Question*: What is the density of A at x where A is a square domain in $I\!\!R^2$, for various points x? (Answer: 0, 1/4, 1/2 or 1.)

PROOF: The theorem is essentially just the Theorem of Section 3.3 applied to the measure $f\lfloor\mathcal{L}$. ∎

Important Example If $A \subset I\!\!R^n$ is measurable then

$$\lim_{r\to 0} \frac{\mathcal{L}(B_r(x)\cap A)}{\mathcal{L}(B_r(x))} = \begin{cases} 1 & \text{a.e. } x \in A \\ 0 & \text{a.e. } x \notin A \end{cases}$$

Definition Assume $f: I\!\!R^n \to I\!\!R$ is locally summable. Then the *precise representative* of f is defined by

$$f^*(x) = \begin{cases} \lim_{r\to 0} \fint_{B_r(x)} f & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}$$

Thus $f^* = f$ a.e. If f is continuous at x, clearly $f^*(x) = f(x)$.

3.5 Riesz Representation Theorem

If μ is a Radon measure on $I\!\!R^n$ then μ induces a positive[37] linear map L on $C_c(I\!\!R^n)$ [38] defined by

$$L(\phi) = \int \phi \, d\mu.$$

Moreover, the converse is true:

Theorem *Assume $L: C_c(I\!\!R^n) \to I\!\!R$ is linear and positive. Then there exists a Radon measure μ on $I\!\!R^n$ such that*

$$L(\phi) = \int \phi \, d\mu$$

for all $\phi \in C_c(I\!\!R^n)$.

---

[37] *Positive* means that $\phi \geq 0$ implies $L(\phi) \geq 0$.

[38] $C_c(I\!\!R^n)$ is the set of compactly supported continuous real valued functions defined on $I\!\!R^n$.

PROOF: The idea of the proof is to first define $\mu(U)$ for open U by

$$\mu(U) = \sup\{L(\phi) : \phi \in C_c(I\!R^n),\ \mathrm{spt}(\phi) \subset U,\ 0 \leq \phi \leq 1\}.$$

Then define

$$\mu(E) = \inf\{\mu(U) : E \subset U,\ U \text{ open}\}.$$

∎

A similar result holds if we drop the positivity requirement:

Theorem *Assume $L : C_c(I\!R^n) \to I\!R$ is linear and*

$$\sup\{L(\phi) : \phi \in C_c(I\!R^n),\ spt(\phi) \subset K,\ |\phi| \leq 1\} < \infty.$$

for each compact $K \subset I\!R^n$. Then there exist mutually singular Radon measures μ^+ and μ^- such that

$$L(\phi) = \int \phi \, d\mu^+ - \int \phi \, d\mu^-.$$

for all $\phi \in C_c(I\!R^n)$.

4 Some Function Spaces

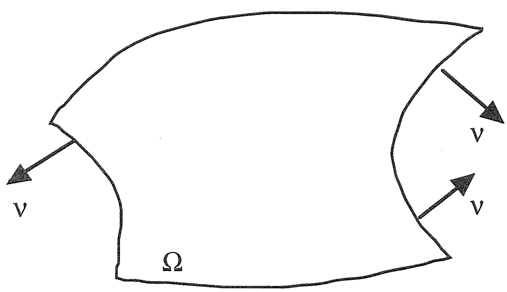
We work with Lebesgue measure in $\mathbb{R}^n$, although many of the results generalise. We assume that $U \subset \mathbb{R}^n$ is open.

4.1 Background Material

4.1.1 Integration by Parts

Suppose Ω is a bounded open subset of $\mathbb{R}^n$ with C^1 boundary,[39] ν is the outward pointing unit normal and $u:\Omega \to \mathbb{R}$ is $C^1(\overline{\Omega})$.[40] Then

1. $\int_\Omega D_i u = \int_{\partial\Omega} \nu_i u$,[41]
2. $\int_\Omega u\, D_i v = -\int_\Omega v\, D_i u + \int_{\partial\Omega} u\nu_i v$.



The first result is the *Divergence Theorem* or *Gauss Green Theorem*. The second follows by replacing u by uv. Note that if u has compact support in Ω, then the boundary terms disappear.

[39]More generally, a finite number of "corners" may be allowed; in fact the boundary may be locally the graph of a Lipschitz function.

[40]$C^1(\overline{\Omega})$ is defined to be the set of *uniformly* continuous functions in $C(\Omega)$. Since Ω is bounded, this is equivalent to the set of continuous functions on the closure of Ω.

[41]Integration over the boundary has not been defined yet. This can be done by integrating with respect to Hausdorff measure $\mathcal{H}^{n-1}$ restricted to the boundary. Equivalently, if the boundary is represented locally as the graph of a function $\phi: U\, (\subset \mathbb{R}^{n-1}) \to \mathbb{R}^n$ and $f:\phi[U]\,(\subset \partial\Omega) \to \mathbb{R}$, then

$$\int_{\phi[U]} f = \int_U f(\phi(x))\, J\phi(x),$$

The *Jacobian* $J\phi(x)$ is here the square root of the sum of the squares of the $(n-1)\times(n-1)$ minors of the $n\times(n-1)$ matrix $D\phi(x)$. Compare this with the Change of Variable Formula in Section 3.2. Note that in the case of domains in $\mathbb{R}^2$, $J\phi(x) = |\nabla\phi(x)|$.

4.1.2 Algebraic Inequalities

Cauchy's Inequality If $a, b \in \mathbb{R}$ then

$$ab \leq \frac{a^2}{2} + \frac{b^2}{2}$$

More generally if $\epsilon > 0$ then

$$ab \leq \epsilon a^2 + \frac{b^2}{4\epsilon}.$$

The point in the second inequality is that "one can dominate ab by a little bit of a^2 at the cost of a lot of b^2".

Young's Inequality If $a, b > 0$, $p > 1$ and $1/p + 1/p' = 1$ then

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

More generally, if also $\epsilon > 0$ then

$$ab \leq \epsilon a^p + c(\epsilon) b^{p'}.$$

In fact, $c(\epsilon) = (p\epsilon)^{-p'/p}/p'$. Note that $c(\epsilon) \uparrow \infty$ as $\epsilon \downarrow 0$.

4.1.3 Integral Inequalities

Suppose $E \subset \mathbb{R}^n$.[42]

Then (*Hölder's inequality*)

$$\int_E |fg| \leq \left(\int_E |f|^p\right)^{1/p} \left(\int_E |g|^q\right)^{1/q}$$

for $p, q > 1$, $1/p + 1/q = 1$.

In particular (*Schwartz's inequality*)

$$\int_E |fg| \leq \left(\int_E |f|^2\right)^{1/2} \left(\int_E |g|^2\right)^{1/2}.$$

Also (*Minkowski's inequality*)

$$\left(\int_E |f+g|^p\right)^{1/p} \leq \left(\int_E |f|^p\right)^{1/p} + \left(\int_E |g|^p\right)^{1/p}.$$

[42] As usual, all sets and functions are assumed measurable.

4.2 L^p Spaces

4.2.1 Definitions

Recall $U \subset I\!R^n$ is open (although this is not necessary for much of what follows).

We define[43] [44]

$$\begin{aligned} \|f\|_{L^p(U)} &= \left(\int_U |f|^p d\mu\right)^{1/p} \quad 1 \le p < \infty, \\ \|f\|_{L^\infty(U)} &= \text{ess sup}_U |f|. \end{aligned}$$

If we "identify" f_1 and f_2 whenever $f_1 = f_2$ a.e. (more precisely, take equivalence classes) and for $1 \le p \le \infty$ define

$$L^p = L^p(U) = \left\{ f : U \to I\!R : \|f\|_{L^p(U)} < \infty \right\},$$

then *$L^p(U)$ is a Banach space with norm $\|\cdot\|_{L^p(U)}$.*[45] Moreover, *$L^2(U)$ is a Hilbert space*[46] *with inner product $(f,g)_{L^2(U)} = \int_U fg$.*

We say

$$f_i \to f \quad \text{in } L^p(U)$$

if $\|f_i - f\|_{L^p(U)} \to 0$ as $i \to \infty$.

Example Suppose $E \subset I\!R^n$ is bounded. Then $1/|x|^\alpha \in L^1(E)$ iff $\alpha < n$. Hence $1/|x|^\alpha \in L^p(E)$ iff $\alpha < n/p$.

Comparing Different L^p Spaces If $|U| < \infty$ [47] then (*Exercise*, using Hölder's inequality)

$$\left(⨍_U |u|^{p_1}\right)^{1/p_1} \le \left(⨍_U |u|^{p_2}\right)^{1/p_2} \quad 1 \le p_1 \le p_2 < \infty,$$

where the *integral average* of f is defined by

$$⨍ f = \frac{1}{|U|}\int f.$$

It follows that if $|U| < \infty$ and $1 \le p_1 \le p_2 \le \infty$ then

$$L^{p_2}(U) \subset L^{p_1}(U).$$

---

[43] By definition, ess $\sup_U |f|$ is the least α such that $|f| \le \alpha$ a.e. in U. It is easy to show that a *least* such α exists, possible $+\infty$.

[44] Note that $\int_U f := \int f\chi_U$ depends only on the values of $f(x)$ for $x \in U$. Instead of extending f to $I\!R^n$ and integrating $\chi_U f$, we could equivalently integrate the "original" f with respect to the measure μ on U defined by restricting $\mathcal{L}$ to subsets of U. Both ways are easily checked to be equivalent.

[45] A *Banach space* is a normed space which is *complete*. The fact $L^p(U)$ is a normed space follows from Minkowski's inequality. Completeness means that if $(f_i)_{i=1}^\infty$ is Cauchy in the L^p norm then $\|f_i - f\|_{L^p} \to 0$ for some $f \in L^p(U)$.

[46] Recall that a Hilbert space is just a Banach space whose norm is given by an inner product.

[47] We often write $|U|$ for $\mathcal{L}(U)$.

Local L^p Spaces We say $f \in L^p_{\text{loc}}(U)$ if $f \in L^p(V)$ for every open $V \subset\subset U$.[48] By "loc" one means "locally". Note that there is no control on $f(x)$ as $x \to \partial U$. Trivially, $L^p(U) \subset L^p_{\text{loc}}(U)$.

The local spaces are *not* normed spaces. We say

$$f_i \to f \quad \text{in } L^p_{\text{loc}}(U)$$

if $\|f_i - f\|_{L^p(V)} \to 0$ as $i \to \infty$ for every open $V \subset\subset U$.

4.2.2 Dual Spaces

It follows from Hölder's inequality that if $1 \leq p \leq \infty$ and $1/p + 1/p' = 1$ then every $f \in L^{p'}(U)$ defines a bounded linear operator F on $L^p(U)$ given by

$$F(g) = \int_U fg.$$

Moreover, $\|F\| = \|f\|_{L^{p'}(U)}$, where $\|F\|$ is the operator norm[49] of F. *Proof*: "$\leq$" follows from Hölder's inequality and "$\geq$" follows from choosing $g = |f|^{p/p'}\operatorname{sign} g$.

If $p \neq \infty$ then all bounded linear operators on L^p are obtained in this manner.

Riesz Representation Theorem *If $1 \leq p < \infty$ then the above map $f \mapsto F$ is an isomorphism from $L^{p'}(U)$ onto the space of bounded linear operators on $L^p(U)$.*

PROOF: If F is a bounded linear operator on $L^p(U)$ then one can apply the second Theorem in Section 3.5 to represent F as a Radon measure μ. One then shows μ is absolutely continuous with respect to Lebesgue measure and so $F = f\lfloor\mathcal{L}$ for some f, c.f. the Radon-Nikodym Theorem in Section 3.3. Finally, one uses Hölder's inequality to deduce $f \in L^{p'}(U)$. ∎

In particular, L^2 is isomorphic to its dual.

There are bounded linear maps on $L^\infty(U)$ which do not correspond to elements of $L^1(U)$. To see this, note that distinct $f \in L^1(U)$ induce distinct bounded linear maps on $C(U) \cap L^\infty(U)$. Since $C(U) \cap L^\infty(U)$ is a *closed* subspace of $L^\infty(U)$, the result follows.

---

[48] $V \subset\subset U$ means the closure of V is a compact subset of U.

[49] That is,

$$\|F\| = \sup\{|F(g)| : \|g\|_{L^p(U)} \leq 1\}.$$

4.2.3 Weak Convergence

Definition Suppose $1 \le p < \infty$. We say $(f_i)_{i=1}^{\infty} \subset L^p(U)$ *converges weakly in* $L^p(U)$ to $f \in L^p(U)$, and write

$$f_i \rightharpoonup f \text{ in } L^p(U),$$

if

$$\int f_i g \to \int fg$$

for all $g \in L^{p'}(U)$.

Remarks

1. Since $L^{p'}(U)$ is the dual of $L^p(U)$ for $1 \le p < \infty$, this is the usual notion of weak convergence for Banach spaces.

2. It is equivalent that (i) $\|f_i\|_{L^p(U)}$ be uniformly bounded, and (ii) $\int f_i g \to \int fg$ for all g in some dense subset S of $L^{p'}(U)$. In particular, if $1 < p < \infty$, one usually takes S to be $C_c(U)$ or $C_c^{\infty}(U)$.

3. If $f_i \to f$ in the usual (strong) sense (i.e. $\|f_i - f\|_{L^p(U)} \to 0$), then $f_i \rightharpoonup f$ in $L^p(U)$, *Exercise*. But the converse is not true.

   The idea for weak convergence is that $f_i \rightharpoonup f$ if f_i converges to f in a sort of "average" sense. For example, let

   $$U = (0,1), \; f_i(x) = \sin(ix), \; f(x) = 0.$$

   Then it follows from *2* that $f_i \rightharpoonup f$, but it is clearly not the case that $f_i \to f$ in $L^p(U)$.

An extremely important fact is that under mild restrictions, if $1 < p < \infty$, then a sequence of functions from $L^p(U)$ will have a subsequence which converges weakly in $L^p(U)$.

Weak Compactness Theorem *Suppose* $1 < p < \infty$. *Suppose* $(f_i)_{i=1}^{\infty} \subset L^p(U)$ *and* $\|f_i\|_{L^p(U)} \le M < \infty$. *Then there exists* $f \in L^p(U)$ *and a subsequence* $f_{i'}$ *such that*

$$f_{i'} \rightharpoonup f \text{ in } L^p(U).$$

PROOF: Since $L^p(U)$ is reflexive[50] for $1 < p < \infty$, this follows from the usual compactness Theorem for the weak* topology. ∎

[50]That is, $L^p(U)$ is isomorphic to the dual of its dual.

The result is not true for strong convergence, as the example in *3* above indicates. Nor is it true if $p = 1$, as we see by taking

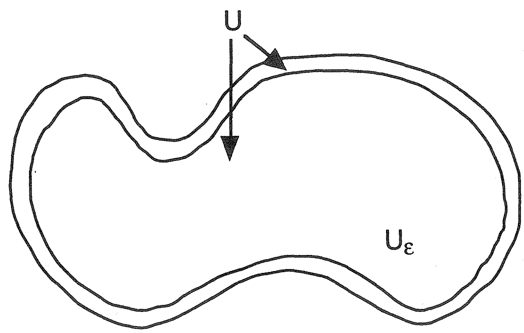
$$U = (-1, 1), \;\; f_i(x) = \begin{cases} ix & 0 \le x \le 1/i \\ 0 & \text{otherwise.} \end{cases}$$

(In fact, this sequence converges "weakly in the sense of measures" to the Dirac measure δ_0.)

4.3 Approximations by Smooth Functions

Define

$$U_\epsilon = \{x \in U : d(x, \partial U)^{51} > \epsilon\}.$$



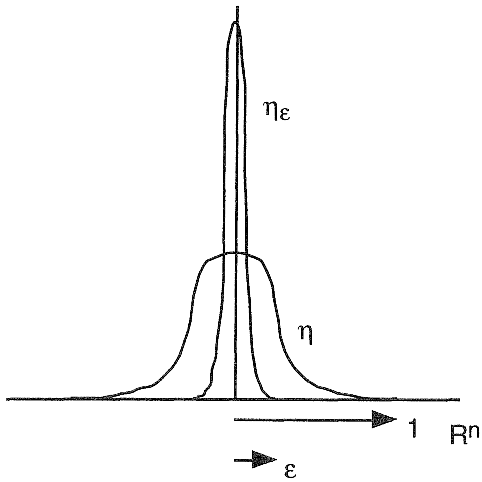
4.3.1 Mollifiers

Fix a C^∞ function $\eta : \mathbb{R}^n \to \mathbb{R}$ such that

1. $\eta \ge 0$,
2. $\eta(x) = 0$ if $|x| \ge 1$,
3. $\int \eta(x)\, dx = 1$.

[51]The *distance* from the point x to the set E is defined by

$$d(x, E) = \inf\{d(x, y) : y \in E\}.$$



Let

$$\eta_\epsilon(x) = \frac{1}{\epsilon^n}\eta\left(\frac{x}{\epsilon}\right).$$

Note that

$$\int \eta_\epsilon = 1, \text{ and } \eta_\epsilon(x) = 0 \text{ if } |x| \geq \epsilon.$$

The function η_ϵ is called a *mollifier.*

For $f \in L^1_{\text{loc}}(U)$ the *ϵ-approximation to f* is defined for $x \in U_\epsilon$ by

$$f_\epsilon(x) = \int \eta_\epsilon(x-y)f(y)\,dy = \int \eta_\epsilon(y)f(x-y)\,dy.[52] \tag{7}$$

We interpret $f_\epsilon(x)$ as a *weighted average* (via η_ϵ) of the values $f(y)$ for y near x.

Suppose f is *uniformly* continuous on U. Then f extends continuously to $\mathbb{R}^n$ by the Tietze Extension Theorem. We often write f_ϵ for the ϵ-approximation to some such extension of f. For $x \in U \setminus U_\epsilon$ the value of $f_\epsilon(x)$ will depend on the particular extension.

Suppose $f \in L^p(U)$. Then the zero extension of f to $\mathbb{R}^n$ belongs to $L^p(\mathbb{R}^n)$. Using this extension we define $f_\epsilon(x)$ for all $x \in U$.

4.3.2 Approximation Results

Theorem *Suppose $U \subset \mathbb{R}^n$ is open and $f \in L^1_{\text{loc}}(U)$. Then*

1. $f_\epsilon \in C^\infty(U_\epsilon)$.
2. *If $f \in C(U)$ then $\sup_{U_\epsilon}|f_\epsilon| \leq \sup_U |f|$ and $f_\epsilon \to f$ uniformly on compact subsets of U.*

[52] The second equality follows from the change of variable formula.

3. If $f \in L^p_{\text{loc}}(U)$ for $1 \leq p < \infty$ then $\|f_\epsilon\|_{L^p(U_\epsilon)} \leq \|f\|_{L^p(U)}$ and $f_\epsilon \to f$ in $L^p_{\text{loc}}(U)$.

4. If $f \in C(U)$ is uniformly continuous, then $f_\epsilon \to f$ uniformly in U.

5. If $f \in L^p(U)$ for $1 \leq p < \infty$ then $f_\epsilon \to f$ in $L^p(U)$.

PROOF: It follows formally by differentiating the *first* expression for f_ϵ in (7) that

$$\frac{\partial f_\epsilon}{\partial x_i} = \int \frac{\partial}{\partial x_i}\eta_\epsilon(x-y)\, f(y)\, dy.$$

This is justified rigorously by taking difference quotients and using the dominated convergence theorem.

One proceeds similarly for higher order derivatives.

The inequality in *2* is easy. To prove uniform convergence, we have by the change of variable formula

$$f_\epsilon(x) = \int_{B_1(0)} \eta(y) f(x - \epsilon y)\, dy.$$

Then *2* and *4* follow, since

$$f_\epsilon(x) - f(x) = \int_{B_1(0)} \eta(y)\left(f(x-\epsilon y) - f(x)\right)\, dy.$$

The inequality in *3* can be established from Hölder's inequality, *Exercise*.

For the remainder of *3* we use the fact that for each $\delta > 0$ and open $V \subset\subset U$, there exists a continuous g such that $\|g - f\|_{L^p(V)} < \delta$.[53] It follows from the inequality in *3* that for ϵ sufficiently small

$$\|g_\epsilon - f_\epsilon\|_{L^p(V)} \leq \|g - f\|_{L^p(U)} < \delta.$$

Using *2*, select ϵ so

$$\|g - g_\epsilon\|_{L^p(V)} \leq \delta.$$

Then *3* follows, and *5* is similar. ∎

Remark Results *3* and *5* are not true if $p = \infty$. A uniform limit of continuous functions is continuous. In particular, if $f(x) = 0$ for $x < 0$ and $f(x) = 1$ for $x \geq 0$, then f is not a limit in the L^∞ norm of continuous functions.

[53] This density result for continuous functions can be established, for example, using Lusin's theorem. Of course it is a weaker case of the result of the present theorem.

4.4 Weak Derivatives

4.4.1 Motivation

In the study of Partial Differential Equations (PDE's), and in the Finite Element Method in Numerical Analysis, to name just two situations, it is necessary to consider functions whose derivatives exist in the so-called *weak sense*. Such functions are called *Sobolev functions*.

For example, it is often fairly easy to show that a (linear or quasilinear) PDE has a Sobolev function as a solution in a certain ("weak") sense. One then attempts to show that such a solution is in fact smooth and is moreover a solution in the classical sense. This can be quite difficult, and will be considered in detail in John Urbas's lectures.

In the finite element method it is necessary to work with continuous and piecewise linear functions. Such functions do not have classical derivatives everywhere, but they *are* Sobolev functions; see Steve Robert's lecture.

Suppose $f \in C^1(U)$, where $U \subset I\!R^n$ is open. Then for all $\phi \in C^1_c(U)$

$$\int D_i f\, \phi = -\int f\, D_i\phi, \quad i = 1, \ldots, n.$$

Moreover, this uniquely determines $D_i f$ in the sense that if

$$\int g\phi = -\int f\, D_i\phi$$

for all $\phi \in C^1_c(U)$, then $g = D_i f$ a.e. *Exercise.*

4.4.2 Introduction

Motivated by the previous considerations, one makes the following definition.

Definition Suppose $f \in L^1_{\mathrm{loc}}(U)$ and $1 \le i \le n$. Then $g_i \in L^1_{\mathrm{loc}}(U)$ is said to be the i^{th} *weak partial derivative* of f in U if

$$\int g_i\phi = -\int f\, D_i\phi$$

for all $\phi \in C^1_c(U)$.

The function g_i is unique (a.e.)[54] and is written $D_i f$ or $\frac{\partial f}{\partial x_i}$. We also write $Df = (D_1 f, \ldots, D_n f)$.

Example 1 If $f \in C^1(U)$ then g_i is the usual (classical) derivative of f.

[54] That is, any two such functions g agree a.e., *Exercise.*

Example 2 Suppose

$$f(x) = |x|.$$

Then the weak derivative exists and is given by

$$Df(x) = \begin{cases} -1 & x \leq 0 \\ 1 & x > 0 \end{cases} \tag{8}$$

To see this, let g be defined by the right side of (8). Then using integration by parts (c.f. Section 4.1.1) we obtain

$$\begin{aligned} \int f\, D_i\phi &= -\int_{-\infty}^{0} x\, D_i\phi + \int_{0}^{\infty} x\, D_i\phi \\ &= \int_{-\infty}^{0} D_i x\, \phi - \int_{0}^{\infty} D_i x\, \phi \\ &= -\int g\phi. \end{aligned}$$

Thus g is indeed the weak derivative.[55] Note in the previous calculation that the boundary terms obtained in passing from the first line to the second line were both zero (more typically, boundary terms will cancel one another).

Example 3 Suppose

$$f(x) = \begin{cases} -1 & x \leq 0 \\ 1 & x > 0 \end{cases}$$

Then the classical derivative exists for all x except $x = 0$, but the weak derivative does *not* exist in the sense of the previous definition. We remark that the weak derivative *does* exist in the *distributional sense* (which is an extension of the present notion of weak derivative), see the lectures by Tony Dooley.

To see this, computing as in the previous example,

$$\begin{aligned} \int f\, D_i\phi &= -\int_{-\infty}^{0} D_i\phi + \int_{0}^{\infty} D_i\phi \\ &= -2\phi(0). \end{aligned}$$

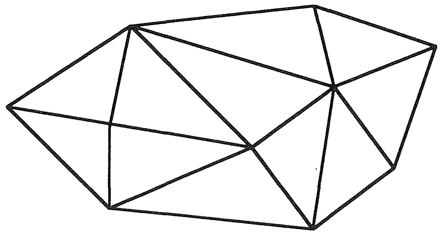
Note that if δ_0 is the Dirac measure at zero (see Section 1.4.2), then it is easy to check that

$$\int \phi\, d\delta_0 = \phi(0)$$

for all $\phi \in C_c(U)$. Thus we can naturally identify the weak derivative of f with $2\delta_0$, but not with any L^1_{loc} function.

[55] Of course, we could have taken $g(0) = 1$, or even have changed g on any set of measure zero.

Example 4 Suppose f is *continuous*, and C^1 on the interior of each triangle, in a triangulated domain as shown below.



Then it is straightforward to check that the two weak derivatives exist, and equal the classical derivatives on the interior of each triangle. The main point is that in performing the integration by parts as in Example 2, one obtains boundary terms which cancel in pairs corresponding to pairs of adjacent triangles.

4.4.3 $W^{1,p}$ Spaces

Definition Let $1 \leq p \leq \infty$. The corresponding *Sobolev Space* and *local Sobolev Space* are defined by

$$\begin{aligned} W^{1,p}(U) &= \{f \in L^p(U) : D_i f \in L^p(U),\ i = 1, \ldots, n\}, \\ W^{1,p}_{\mathrm{loc}}(U) &= \{f \in L^p_{\mathrm{loc}}(U) : D_i f \in L^p_{\mathrm{loc}}(U),\ i = 1, \ldots, n\}, \end{aligned}$$

where the $D_i f$ is the weak derivatives. The $W^{1,p}$ norm is defined by

$$\begin{aligned} \|f\|_{W^{1,p}(U)} &= \left(\int_U |f|^p + \sum_{i=1}^n |D_i f|^p \right)^{1/p} \quad 1 \leq p < \infty, \\ \|f\|_{W^{1,\infty}(U)} &= \text{ess sup}_U \left(|f| + \sum_{i=1}^n |D_i f| \right). \end{aligned}$$

Remark $W^{1,p}(U)$ together with the norm $\|\cdot\|_{W^{1,p}(U)}$ is a Banach space.
Proof: The main point is that $W^{1,p}(U)$ is a closed subspace of the $n+1$-fold product Banach space $L^p(U) \times \cdots \times L^p(U)$.

We write

$$f_i \to f \text{ in } W^{1,p}(U)$$

if $\|f_i - f\|_{W^{1,p}(U)} \to 0$, and

$$f_i \to f \text{ in } W^{1,p}_{\mathrm{loc}}(U)$$

if $\|f_i - f\|_{W^{1,p}(V)} \to 0$ for each open $V \subset\subset U$.

Theorem *If the weak derivatives $D_i f$ of f exist, then*

$$D_i(f_\epsilon) = (D_i f)_\epsilon.$$

Moreover, if $f \in W^{1,p}_{\text{loc}}(U)$ for some $1 \leq p < \infty$, then

$$f_\epsilon \to f \text{ in } W^{1,p}_{\text{loc}}(U).$$

PROOF: One computes

$$\begin{aligned} D_i(f_\epsilon) &= \int \frac{\partial \eta_\epsilon}{\partial x_i}(x-y) f(y)\, dy \\ &= -\int \frac{\partial \eta_\epsilon}{\partial y_i}(x-y) f(y)\, dy \\ &= \int \eta_\epsilon(x-y) \frac{\partial f}{\partial y_i}(y)\, dy \\ &= (D_i f)_\epsilon. \end{aligned}$$

The convergence follows from the Theorem in Section 4.3.2. ∎

For further important properties of Sobolev Spaces see [EG].

References

[EG] L. Evans & R. Gariepy, *Measure Theory and Fine Properties of Functions,* CRC Press 1992.

[R] H.L. Royden, *Real Analysis,* Macmillan 2nd ed. 1968, 3rd ed. 1988.

[S] K.T. Smith, *Primer of Modern Analysis,* Bogden & Quigley 1971, Springer.