JUHA KINNUNEN

Real Analysis

Department of Mathematics and Systems Analysis, Aalto University
2020

Contents

The L^p spaces are probably the most important function spaces in analysis. This section gives basic facts about L^p spaces for general measures. These include Hölder's inequality, Minkowski's inequality, the Riesz-Fischer theorem which shows the completeness and the corresponding facts for the L^∞ space.

1 L^p spaces

In this section we study the L^p spaces in order to be able to capture finer quantitative information on the size of measurable functions and effect of operators on such functions. The cases $0 < p < 1$, $p = 1$, $p = 2$, $1 < p < \infty$ and $p = \infty$ are different in character, but they are all play an important role in modern analysis, for example, in Fourier and harmonic analysis, functional analysis and partial differential equations. The space L^1 of integrable functions plays a central role in measure and integration theory. The Hilbert space L^2 of square integrable functions is important in the study of Fourier series. Many operators that arise in applications are bounded in L^p for $1 < p < \infty$, but the limit cases L^1 and L^∞ require a special attention.

1.1 L^p functions

Definition 1.1. Let μ be an outer measure on $\mathbb{R}^n$, $A \subset \mathbb{R}^n$ a μ-measurable set and $f : \mathbb{R}^n \to [-\infty, \infty]$ a μ-measurable function. Then $f \in L^p(A)$, $1 \leqslant p < \infty$, if

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

THE MORAL: For $p = 1$, $f \in L^1(A)$ if and only if $|f|$ is integrable in A. For $1 \leqslant p < \infty$, $f \in L^p(A)$ if and only if $|f|^p$ is integrable in A.

Remark 1.2. The measurability assumption on f essential in the definition. For example, let $A \subset [0,1]$ be a non-measurable set with respect to the one-dimensional Lebesgue measure and consider $f : [0,1] \to \mathbb{R}$,

$$f(x) = \begin{cases} 1, & x \in A, \\ -1, & x \in [0,1] \setminus A. \end{cases}$$

Then $f^2 = 1$ is integrable on $[0,1]$, but f is not a Lebesgue measurable function.

Example 1.3. Let $f : \mathbb{R}^n \to [0,\infty]$, $f(x) = |x|^{-n}$ and assume that μ is the Lebesgue measure. Let $A = B(0,1) = \{x \in \mathbb{R}^n : |x| < 1\}$ and denote $A_i = B(0,2^{-i+1}) \setminus B(0,2^{-i})$, $i = 1,2,\dots$. Then

$$\begin{aligned}
\int_{B(0,1)} |x|^{-np}\,dx &= \sum_{i=1}^{\infty} \int_{A_i} |x|^{-np}\,dx \\
&\leqslant \sum_{i=1}^{\infty} \int_{A_i} 2^{npi}\,dx \qquad (x \in A_i \Rightarrow |x| \geqslant 2^{-i} \Rightarrow |x|^{-np} \leqslant 2^{npi}) \\
&= \sum_{i=1}^{\infty} 2^{npi}|A_i| \leqslant \sum_{i=1}^{\infty} 2^{npi}|B(0,2^{-i+1})| \\
&= \Omega_n \sum_{i=1}^{\infty} 2^{npi}(2^{-i+1})^n \qquad (\Omega_n = |B(0,1)|) \\
&= \Omega_n \sum_{i=1}^{\infty} 2^{npi-ni+n} = 2^n \Omega_n \sum_{i=1}^{\infty} 2^{in(p-1)} < \infty,
\end{aligned}$$

if $n(p-1) < 0 \iff p < 1$. Thus $f \in L^p(B(0,1))$, if $p < 1$.

On the other hand,

$$\begin{aligned}
\int_{B(0,1)} |x|^{-np}\,dx &= \sum_{i=1}^{\infty} \int_{A_i} |x|^{-np}\,dx \\
&\geqslant \sum_{i=1}^{\infty} \int_{A_i} 2^{np(i-1)}\,dx \qquad (x \in A_i \Rightarrow |x| < 2^{-i+1} \Rightarrow |x|^{-np} > 2^{np(i-1)}) \\
&= \sum_{i=1}^{\infty} 2^{np(i-1)}|A_i| = \Omega_n(2^n-1)2^{-np} \sum_{i=1}^{\infty} 2^{npi}2^{-in} \\
&\qquad\qquad (|A_i| = |B(0,2^{-i+1})| - |B(0,2^{-i})| \\
&\qquad\qquad\quad = \Omega_n(2^{(-i+1)n} - 2^{-in}) = \Omega_n(2^n-1)2^{-in}) \\
&= C(n,p) \sum_{i=1}^{\infty} 2^{in(p-1)} = \infty,
\end{aligned}$$

if $n(p-1) \geqslant 0 \iff p \geqslant 1$. Thus $f \notin L^p(B(0,1))$, if $p \geqslant 1$. This shows that

$$f \in L^p(B(0,1)) \iff p < 1.$$

If $A = \mathbb{R}^n \setminus B(0,1)$, then we denote $A_i = B(0,2^i) \setminus B(0,2^{i-1})$, $i = 1,2,\dots$, and a similar argument as above shows that

$$f \in L^p(\mathbb{R}^n \setminus B(0,1)) \iff p > 1.$$

Observe that $f \notin L^1(B(0,1))$ and $f \notin L^1(\mathbb{R}^n \setminus B(0,1))$. Thus $f(x) = |x|^{-n}$ is a borderline function in $\mathbb{R}^n$ as far as integrability is concerned.

T H E M O R A L : The smaller the parameter p is, the worse local singularities an L^p function may have. On the other hand, the larger the parameter p is, the more an L^p function may spread out globally.

Example 1.4. Suppose that $f : \mathbb{R}^n \to [0,\infty]$ is radial. Thus f depends only on $|x|$ and it can be expressed as $f(|x|)$, where f is a function defined on $[0,\infty)$. Then

$$\int_{\mathbb{R}^n} f(|x|)\,dx = \omega_{n-1}\int_0^\infty f(r)r^{n-1}\,dr, \tag{1.1}$$

where

$$\omega_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

is the $(n-1)$-dimensional volume of the unit sphere $\partial B(0,1) = \{x \in \mathbb{R}^n : |x| = 1\}$.

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, we have

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

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

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

Let $r > 0$. Then

$$\begin{aligned} \int_{\mathbb{R}^n\setminus B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \int_{\mathbb{R}^n} \frac{1}{|x|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,r)}(x)\,dx \\ &= r^n\int_{\mathbb{R}^n} \frac{1}{|rx|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,r)}(rx)\,dx \\ &= r^{n-\alpha}\int_{\mathbb{R}^n} \frac{1}{|x|^\alpha}\chi_{\mathbb{R}^n\setminus B(0,1)}(x)\,dx \\ &= r^{n-\alpha}\int_{\mathbb{R}^n\setminus B(0,1)} \frac{1}{|x|^\alpha}\,dx < \infty, \quad \alpha > n, \end{aligned}$$

and, in a similar way,

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

Observe, that here we formally make the change of variables $x = ry$.

On the other hand, the integrals can be computer directly by (1.1). This gives

$$\begin{aligned} \int_{\mathbb{R}^n\setminus B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \omega_{n-1}\int_r^\infty \rho^{-\alpha}\rho^{n-1}\,d\rho \\ &= \frac{\omega_{n-1}}{-\alpha+n}\rho^{-\alpha+n}\Big|_r^\infty = \frac{\omega_{n-1}}{\alpha-n}r^{-\alpha+n} < \infty, \quad \alpha > n \end{aligned}$$

and

$$\begin{aligned} \int_{B(0,r)} \frac{1}{|x|^\alpha}\,dx &= \omega_{n-1}\int_0^r \rho^{-\alpha}\rho^{n-1}\,d\rho \\ &= \frac{\omega_{n-1}}{-\alpha+n}\rho^{-\alpha+n}\Big|_0^r = \frac{\omega_{n-1}}{\alpha-n}r^{n-\alpha} < \infty, \quad \alpha < n. \end{aligned}$$

Remarks 1.5:
Formula (1.1) implies following claims:

(1) If $|f(x)| \leqslant c|x|^{-\alpha}$ in a ball $B(0,r)$, $r>0$, for some $\alpha<n$, then $f \in L^1(B(0,r))$. On the other hand, if $|f(x)| \geqslant c|x|^{-\alpha}$ in $B(0,r)$ for some $\alpha>n$, then $f \notin L^1(B(0,r))$.

(2) If $|f(x)| \leqslant c|x|^{-\alpha}$ in $\mathbb{R}^n \setminus B(0,r)$ for some $\alpha>n$, then $f \in L^1(\mathbb{R}^n \setminus B(0,r))$. On the other hand, if $|f(x)| \geqslant c|x|^{-\alpha}$ in $\mathbb{R}^n \setminus B(0,r)$ for some $\alpha<n$, then $f \notin L^1(\mathbb{R}^n \setminus B(0,r))$.

Remark 1.6. $f \in L^p(A) \Longrightarrow |f(x)|<\infty$ for μ-almost every $x \in A$.

Reason. Let $A_i=\{x \in A : |f(x)| \geqslant i\}$, $i=1,2,\dots$. Then

$$\{x \in A : |f(x)|=\infty\} = \bigcap_{i=1}^{\infty} A_i$$

and

$$\begin{aligned}
\{x \in A : |f(x)|=\infty\} \leqslant \mu(A_i) &= \int_{A_i} 1\,d\mu \\
&\leqslant \int_{A_i} \left(\frac{|f|}{i}\right)^p d\mu \qquad (|f| \geqslant i \text{ in } A_i) \\
&\leqslant \frac{1}{i^p} \underbrace{\int_A |f|^p\,d\mu}_{<\infty} \to 0 \quad \text{as} \quad i \to \infty.
\end{aligned}$$

■

The converse is not true, as the previous example shows.

1.2 L^p norm

If $f \in L^p(A)$, $1 \leqslant p < \infty$, the norm of f is the number

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

We shall see that this has the usual properties of the norm:

(1) (Nonnegativity) $0 \leqslant \|f\|_p < \infty$,

(2) $\|f\|_p = 0 \Longleftrightarrow f=0$ μ-almost everywhere,

(3) (Homogeneity) $\|af\|_p = |a|\|f\|_p$, $a \in \mathbb{R}$,

(4) (Triangle inequality) $\|f+g\|_p \leqslant \|f\|_p + \|g\|_p$.

The claims (1) and (3) are clear. For $p=1$, the claim (4) follows from the pointwise triangle inequality $|f(x)+g(x)| \leqslant |f(x)|+|g(x)|$. For $p>1$, the claim (4) is not trivial and we shall prove it later in this section.

Let us recall how to prove (2). Recall that if a property holds except on a set of μ measure zero, we say that it holds μ-almost everywhere.

$\boxed{\Leftarrow}$: Assume that $f = 0$ μ-almost everywhere in A. Then

$$\int_A |f|^p \, d\mu = \underbrace{\int_{A\cap\{|f|=0\}} |f|^p \, d\mu}_{\substack{=0, \\ |f|=0 \ \mu\text{-a.e.}}} + \underbrace{\int_{A\cap\{|f|>0\}} |f|^p \, d\mu}_{\substack{=0, \\ \mu(A\cap\{|f|>0\})=0}} = 0.$$

Thus $\|f\|_p = 0$.

$\boxed{\Rightarrow}$: Assume that $\|f\|_p = 0$. Let $A_i = \left\{x \in A : |f(x)| \geqslant \frac{1}{i}\right\}$, $i = 1,2,\ldots$. Then

$$\{x \in A : |f(x)| > 0\} = \bigcup_{i=1}^{\infty} A_i$$

and

$$\mu(A_i) = \int_{A_i} 1 \, d\mu \leqslant \int_{A_i} |if|^p \, d\mu \leqslant i^p \underbrace{\int_A |f|^p \, d\mu}_{=0} = 0. \qquad (i|f| \geqslant 1 \text{ in } A_i)$$

Thus $\mu(A_i) = 0$ for every $i = 1,2,\ldots$ and

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mu(A_i) = 0.$$

In other words, $f = 0$ μ-almost everywhere in A.

If f and g are two μ-measurable functions on a μ-measurable set A, then we shall be very much interested in the case $f(x) = g(x)$ for μ-almost every $x \in A$. Of course, this means that $\mu(\{x \in A : f(x) \neq g(x)\}) = 0$. In the case $f = g$ μ-almost everywhere, we do not usually distinguish f from g. That is, we shall regard them as equal. We could be very formal and introduce the equivalence relation

$$f \sim g \Longleftrightarrow f = g \quad \mu\text{-almost everywhere in } A$$

but this is hardly necessary. In practice, we are thinking f as the equivalence class of all functions which are equal to f μ-almost everywhere in A. Thus $L^p(A)$ actually consists of equivalence classes rather than functions, but we shall not make the distinction. Indeed, in measure and integration theory we cannot distinguish f from g, if the functions are equal μ-almost everywhere. In fact, if $f = g$ μ-almost everywhere in A, then $f \in L^p(A) \Longleftrightarrow g \in L^p(A)$ and $\|f - g\|_p = 0$. In particular, this implies that $\|f\|_p = \|g\|_p$. On the other hand, if $\|f - g\|_p = 0$, then $f = g$ μ-almost everywhere in A.

Another situation that frequently arises is that the function f is defined only almost everywhere. Then we say that f is measurable if and only if its zero extension to the whole space is measurable. Observe, that this does not affect the L^p norm of f.

Next we show that $L^p(A)$ is a vector space.

Lemma 1.7. (i) If $f \in L^p(A)$, then $af \in L^p(A)$, $a \in \mathbb{R}$.

(ii) If $f, g \in L^p(A)$, then $f+g \in L^p(A)$.

Proof. $\boxed{(1)}$:

$$\int_A |af|^p \, d\mu = |a|^p \int_A |f|^p \, d\mu < \infty.$$

$\boxed{(2)}$ $\boxed{p=1}$: The triangle inequality $|f+g| \leqslant |f|+|g|$ implies

$$\int_A |f+g| \, d\mu \leqslant \int_A |f| \, d\mu + \int_A |g| \, d\mu < \infty.$$

$\boxed{1 < p < \infty}$: The elementary inequality

$$\begin{aligned}(a+b)^p &\leqslant (a+b)^p \leqslant (2\max(a,b))^p \\ &= 2^p \max(a^p, b^p) \leqslant 2^p(a^p+b^p), \quad a, b \in \mathbb{R}, \quad 0 < p < \infty \end{aligned} \tag{1.2}$$

implies

$$\int_A |f+g|^p \, d\mu \leqslant 2^p \left(\int_A |f|^p \, d\mu + \int_A |g|^p \, d\mu \right) < \infty.$$ □

Remark 1.8. Note that the proof applies for $0 < p < \infty$. Thus $L^p(A)$ is a vector space for $0 < p < \infty$. However, it will be a normed space only for $p \geqslant 1$ as we shall see later.

Remark 1.9. A more careful analysis gives the useful inequality

$$(a+b)^p \leqslant 2^{p-1}(a^p+b^p), \quad a, b \in \mathbb{R}, \quad 1 \leqslant p < \infty. \tag{1.3}$$

Remarks 1.10:

(1) If $f : A \to \mathbb{C}$ is a complex-valued function, then f is said to be μ-measurable if and only if $\operatorname{Re} f$ and $\operatorname{Im} f$ are μ-measurable. We say that $f \in L^1(A)$ if $\operatorname{Re} f \in L^1(A)$ and $\operatorname{Im} f \in L^1(A)$, and we define

$$\int_A f \, d\mu = \int_A \operatorname{Re} f \, d\mu + i \int_A \operatorname{Im} f \, d\mu,$$

where i is the imaginary unit. This integral satisfies the usual linearity properties. It also satisfies the important inequality

$$\left| \int_A f \, d\mu \right| \leqslant \int_A |f| \, d\mu.$$

The definition of the L^p spaces and the norm extends in a natural way to complex-valued functions. Note that the property $\|af\|_p = |a| \|f\|_p$ for every $a \in \mathbb{C}$ and thus L^p is a complex vector space.

(2) The space $L^2(A)$ is an inner product space with the inner product

$$\langle f,g\rangle = \int_A f\overline{g}\,d\mu, \quad f,g \in L^2(A).$$

Here $\overline{g}$ is the complex conjugate which can be neglected if the functions are real-valued. This inner product induces the standard L^2-norm, since

$$\|f\|_2 = \left(\int_A |f|^2\,d\mu\right)^{1/2} = \left(\int_A f\overline{f}\,d\mu\right)^{1/2} = \langle f,f\rangle^{1/2}.$$

(3) In the special case that $A = \mathbb{N}$ and μ is the counting measure, the $L^p(\mathbb{N})$ spaces are denoted by l^p and

$$l^p = \left\{(x_i) : \sum_{i=1}^{\infty} |x_i|^p < \infty\right\}, \quad 1 \leqslant p < \infty.$$

Here (x_i) is a sequence of real (or complex) numbers. In this case,

$$\int_{\mathbb{N}} x\,d\mu = \sum_{i=1}^{\infty} x(i)$$

for every nonnegative function x on $\mathbb{N}$. Thus

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p}.$$

Note that the theory of L^p spaces applies to these sequence spaces as well.

Definition 1.11. Let $1 < p < \infty$. The Hölder conjugate p' of p is the number which satisfies

$$\frac{1}{p} + \frac{1}{p'} = 1.$$

For $p = 1$ we define $p' = \infty$ and if $p = \infty$, then $p' = 1$.

Remark 1.12. Note that

$$\begin{aligned}
&p' = \frac{p}{p-1},\\
&p = 2 \Longrightarrow p' = 2,\\
&1 < p < 2 \Longrightarrow p' > 2,\\
&2 < p < \infty \Longrightarrow 1 < p' < 2,\\
&p \to 1 \Longrightarrow p' \to \infty,\\
&(p')' = p.
\end{aligned}$$

Lemma 1.13 (Young's inequality). Let $1 < p < \infty$. Then for every $a \geqslant 0$, $b \geqslant 0$,

$$ab \leqslant \frac{a^p}{p} + \frac{b^{p'}}{p'},$$

with equality if and only if $a^p = b^{p'}$.

T H E M O R A L : Young's inequality is a very useful tool in splitting a product to a sum. Morever, it shows where the conjugate exponent p' comes from.

Proof. The claim is obviously true, if $a = 0$ or $b = 0$. Thus we may assume that $a > 0$ and $b > 0$. Clearly

$$ab \leqslant \frac{a^p}{p} + \frac{b^{p'}}{p'} \iff \frac{1}{p}\frac{a^p}{b^{p'}} + \frac{1}{p'} - ab^{1-p'} \geqslant 0 \iff \frac{1}{p}\left(\frac{a}{b^{p'/p}}\right)^p + \frac{1}{p'} - \frac{a}{b^{p'/p}} \geqslant 0$$

Let $t = a/b^{p'/p}$ and define $\varphi : (0,\infty) \to \mathbb{R}$,

$$\varphi(t) = \frac{1}{p}t^p + \frac{1}{p'} - t.$$

Then

$$\varphi(0) = \frac{1}{p'}, \quad \lim_{t\to\infty}\varphi(t) = \infty \quad \text{and} \quad \varphi'(t) = t^{p-1} - 1.$$

Note that $\varphi'(t) = 0 \iff t = 1$, from which we conclude

$$\varphi(t) \geqslant \varphi(1) = \frac{1}{p} + \frac{1}{p'} - 1 = 0 \quad \text{for every } t > 0.$$

Moreover, $\varphi(t) > 0$, if $t \neq 1$. It follows that $\varphi(t) = 0$ if and only if $a/b^{p'/p} = t = 1$. □

Remarks 1.14:

(1) Young's inequality for $p = 2$ follows immediately from

$$(a-b)^2 \geqslant 0 \iff a^2 - 2ab + b^2 \geqslant 0 \iff \frac{a^2}{2} + \frac{b^2}{2} \geqslant ab \geqslant 0.$$

(2) Young's inequality can be also proved geometrically. To see this, consider the curves $y = x^{p-1}$ and the inverse $x = y^{1/(p-1)} = y^{p'-1}$. Then

$$\int_0^a x^{p-1}\,dx = \frac{a^p}{p} \quad \text{and} \quad \int_0^b y^{p'-1}\,dy = \frac{b^{p'}}{p'}.$$

By comparing the areas under the curves that these integrals measure, we have

$$ab \leqslant \int_0^a x^{p-1}\,dx + \int_0^b y^{p'-1}\,dy = \frac{a^p}{p} + \frac{b^{p'}}{p'}.$$

Theorem 1.15 (Hölder's inequality). Let $1 < p < \infty$ and assume that $f \in L^p(A)$ and $g \in L^{p'}(A)$. Then $fg \in L^1(A)$ and

$$\int_A |fg|\,d\mu \leqslant \left(\int_A |f|^p\,d\mu\right)^{1/p}\left(\int_A |g|^{p'}\,d\mu\right)^{1/p'}.$$

Moreover, an equality occurs if and only if there exists a constant c such that $|f(x)|^p = c|g(x)|^{p'}$ for μ-almost every $x \in A$.

T H E M O R A L : Hölder's inequality is very useful tool in estimating a product of functions.

Remark 1.16. Hölder's inequality states that $\|fg\|_1 \leqslant \|f\|_p \|g\|_{p'}$, $1 < p < \infty$. Observe that for $p = 2$ this is the Cauchy-Schwarz inequality $|\langle f, g\rangle| \leqslant \|f\|_2 \|g\|_2$.

Proof. If $\|f\|_p = 0$, then $f = 0$ μ-almost everywhere in A and thus $fg = 0$ μ-almost everywhere in A. Thus the result is clear, if $\|f\|_p = 0$ or $\|g\|_p = 0$. We can therefore assume that $\|f\|_p > 0$ and $\|g\|_p > 0$. Define

$$\widetilde{f} = \frac{f}{\|f\|_p} \quad \text{and} \quad \widetilde{g} = \frac{g}{\|g\|_{p'}}.$$

Then

$$\|\widetilde{f}\|_p = \left\| \frac{f}{\|f\|_p} \right\|_p = \frac{\|f\|_p}{\|f\|_p} = 1 \quad \text{and} \quad \|\tilde{g}\|_{p'} = 1.$$

By Young's inequality

$$\begin{aligned} \frac{1}{\|f\|_p \|g\|_{p'}} \int_A |fg|\, d\mu &= \int_A |\widetilde{f}\widetilde{g}|\, d\mu \\ &\leqslant \frac{1}{p} \underbrace{\int_A |\widetilde{f}|^p \, d\mu}_{=1} + \frac{1}{p'} \underbrace{\int_A |\widetilde{g}|^{p'} \, d\mu}_{=1} = \frac{1}{p} + \frac{1}{p'} = 1. \end{aligned}$$

An equality holds if and only if

$$\int_A \underbrace{\left(\frac{1}{p} |\widetilde{f}|^p + \frac{1}{p'} |\widetilde{g}|^{p'} - |\widetilde{f}\widetilde{g}| \right)}_{\geqslant 0} d\mu = 0,$$

which implies that

$$\frac{1}{p} |\widetilde{f}|^p + \frac{1}{p'} |\widetilde{g}|^{p'} - |\widetilde{f}\widetilde{g}| = 0 \quad \mu\text{-almost everywhere in } A.$$

The equality occurs in Young's inequality if and only if $|\widetilde{f}|^p = |\widetilde{g}|^{p'}$ μ-almost everywhere in A. Thus

$$|f(x)|^p = \frac{\|f\|_p^p}{\|f\|_{p'}^{p'}} |g(x)|^{p'} \quad \text{for } \mu\text{-almost every } x \in A.$$

□

WARNING: $f \in L^p(A)$ and $g \in L^p(A)$ does not imply that $fg \in L^p(A)$.

Reason. Let

$$f : (0,1) \to \mathbb{R},\ f(x) = \frac{1}{\sqrt{x}},\ g = f,$$

and assume that μ is the Lebesgue measure. Then $f \in L^1((0,1))$ and $g \in L^1((0,1))$, but

$$(fg)(x) = f(x)g(x) = \frac{1}{x} \quad \text{and} \quad fg \notin L^1((0,1)).$$

■

Remarks 1.17:

(1) For $p = 2$ we have Schwarz's inequality

$$\int_A |fg|\,d\mu \leqslant \left(\int_A |f|^2\,d\mu\right)^{1/2} \left(\int_A |g|^2\,d\mu\right)^{1/2}.$$

(2) Hölder's inequality holds for arbitrary measurable functions with the interpretation that the integrals may be infinite. (Exercise)

Lemma 1.18 (Jensen's inequality). Let $1 \leqslant p < q < \infty$ and assume that $0 < \mu(A) < \infty$. Then

$$\left(\frac{1}{\mu(A)}\int_A |f|^p\,d\mu\right)^{1/p} \leqslant \left(\frac{1}{\mu(A)}\int_A |f|^q\,d\mu\right)^{1/q}.$$

T H E M O R A L : The integral average is an increasing function of the power.

Proof. By Hölder's inequality

$$\begin{aligned}\int_A |f|^p\,d\mu &\leqslant \left(\int_A |f|^{pq/p}\,d\mu\right)^{p/q} \left(\int_A 1^{q/(q-p)}\,d\mu\right)^{(q-p)/q}\\ &= \left(\int_A |f|^q\,d\mu\right)^{p/q} \mu(A)^{1-p/q}.\end{aligned}$$

□

Remark 1.19. If $1 \leqslant p < q < \infty$ and $\mu(A) < \infty$, then $L^q(A) \subset L^p(A)$.

W A R N I N G : Let $1 \leqslant p < q < \infty$. In general, $L^q(A) \not\subset L^p(A)$ or $L^p(A) \not\subset L^q(A)$.

Reason. Let $f : (0,\infty) \to \mathbb{R}$, $f(x) = x^a$ and assume that μ is the Lebesgue measure. Then

$$f \in L^1((0,1)) \iff a > -1 \quad \text{and} \quad f \in L^1((1,\infty)) \iff a < -1.$$

Assume that $1 \leqslant p < q < \infty$. Choose b such that $1/q \leqslant b \leqslant 1/p$. Then the function $x^{-b}\chi_{(0,1)}(x)$ belongs to $L^p((0,\infty))$, but does not belong to $L^q((0,\infty))$. On the other hand, the function $x^{-b}\chi_{(1,\infty)}(x)$ belongs to $L^q((0,\infty))$, but does not belong to $L^p((0,\infty))$. ■

Examples 1.20:

(1) Let $A = (0,1)$, μ be the Lebesgue measure and $1 \leqslant p < \infty$. Define $f : (0,1) \to \mathbb{R}$,

$$f(x) = \frac{1}{x^{1/p}(\log(2/x))^{2/p}}.$$

Then $f \in L^p((0,1))$, but $f \notin L^q((0,1))$ for any $q > p$. Thus for every p with $1 \leqslant p < \infty$, there exists a function f which belongs to $L^p((0,1))$, but does not belong to any higher $L^q((0,1))$ with $q > p$. (Exercise)

(2) Let $1 \leqslant p < q < \infty$. Assume that A contains μ-measurable sets of arbitrarily small positive measure. Then there are pairwise disjoint μ-measurable sets $A_i \subset A$, $i = 1,2,\dots$, such that $\mu(A_i) > 0$ and $\mu(A_i) \to 0$ as $i \to \infty$. Let

$$f = \sum_{i=1}^{\infty} a_i \chi_{A_i},$$

where $a_i \to \infty$ are chosen so that

$$\sum_{i=1}^{\infty} a_i^q \mu(A_i) = \infty \quad \text{and} \quad \sum_{i=1}^{\infty} a_i^p \mu(A_i) < \infty.$$

Then $f \in L^p(A) \setminus L^q(A)$. It can be shown, that $L^p(A)$ is not contained in $L^q(A)$ if and only if A contains measurable sets of arbitrarily small positive measure. (Exercise)

(3) Let $1 \leqslant p < q < \infty$. Assume that A contains μ-measurable sets of arbitrarily large measure. Then there are pairwise disjoint μ-measurable sets $A_i \subset A$, $i = 1, 2, \dots$, such that $\mu(A_i) > 0$ and $\mu(A_i) \to \infty$ as $i \to \infty$. Let

$$f = \sum_{i=1}^{\infty} a_i \chi_{A_i},$$

where $a_i \to 0$ are chosen so that

$$\sum_{i=1}^{\infty} a_i^q \mu(A_i) < \infty \quad \text{and} \quad \sum_{i=1}^{\infty} a_i^p \mu(A_i) = \infty.$$

Then $f \in L^q(A) \setminus L^p(A)$. It can be shown, that $L^q(A)$ is not contained in $L^p(A)$ if and only if A contains measurable sets of arbitrarily large measure. (Exercise)

Remark 1.21. There is a more general version of Jensen's inequality. Assume that $0 < \mu(A) < \infty$. Let $f \in L^1(A)$ such that $a < f(x) < b$ for every $x \in A$. If φ is a convex function on (a, b), then

$$\varphi\left(\frac{1}{\mu(A)} \int_A f \, d\mu\right) \leqslant \frac{1}{\mu(A)} \int_A \varphi \circ f \, d\mu.$$

The cases $a = -\infty$ and $b = \infty$ are not excluded. Observe, that in this case may happen that $\varphi \circ f$ is not integrable. We leave the proof as an exercise.

Theorem 1.22 (Minkowski's inequality). Assume $1 \leqslant p < \infty$ and $f, g \in L^p(A)$. Then $f + g \in L^p(A)$ and

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p.$$

Moreover, an equality occurs if and only if there exists a positive constant c such that $f(x) = cg(x)$ for μ-almost every $x \in A$.

THE MORAL: Minkowski's inequality is the triangle inequality for the L^p-norm. It implies that the L^p norm, with $1 \leqslant p < \infty$, is a norm in the usual sense and that $L^p(A)$ is a normed space if the functions that coincide almost everywhere are identified.

Remark 1.23. Elementary inequalities (1.3) and (1.4) imply that

$$\begin{aligned}\|f+g\|_p &= \left(\int_A |f+g|^p\,d\mu\right)^{1/p} \leqslant 2^{(p-1)/p}\left(\int_A (|f|^p+|g|^p)\,d\mu\right)^{1/p}\\ &\leqslant 2^{(p-1)/p}\left(\left(\int_A |f|^p\,d\mu\right)^{1/p}+\left(\int_A |g|^p\,d\mu\right)^{1/p}\right)\\ &= 2^{(p-1)/p}(\|f\|_p+\|g\|_p),\quad 1\leqslant p<\infty.\end{aligned}$$

Observe that the factor $2^{(p-1)/p}$ is strictly greater than one for $p>1$ and Minkowski's inequality does not follow from this.

Proof. $\boxed{p=1}$: The triangle inequality, as in the proof of Lemma 1.7, shows that $\|f+g\|_1 \leqslant \|f\|_1+\|g\|_1$.

$\boxed{1<p<\infty}$: If $\|f+g\|_p=0$, there is nothing to prove. Thus we may assume that $\|f+g\|_p>0$. Then by Hölder's inequality

$$\begin{aligned}\int_A |f+g|^p\,d\mu &\leqslant \int_A |f+g|^{p-1}|f|\,d\mu+\int_A |f+g|^{p-1}|g|\,d\mu\\ &\qquad\qquad (|f+g|^p=|f+g|^{p-1}|f+g|\leqslant |f+g|^{p-1}(|f|+|g|))\\ &\leqslant \left(\int_A |f+g|^{(p-1)p'}\,d\mu\right)^{1/p'}\left(\int_A |f|^p\,d\mu\right)^{1/p}\\ &\quad+\left(\int_A |f+g|^{(p-1)p'}\,d\mu\right)^{1/p'}\left(\int_A |g|^p\,d\mu\right)^{1/p}.\end{aligned}$$

Since $\|f+g\|_p>0$, we have

$$\left(\int_A |f+g|^p\,d\mu\right)^{1-(p-1)/p} \leqslant \left(\int_A |f|^p\,d\mu\right)^{1/p}+\left(\int_A |g|^p\,d\mu\right)^{1/p}.$$

It remains to consider when the equality can occur. This happens if there is an equality in the pointwise inequality $|f+g|^p=|f+g|^{p-1}|f+g|\leqslant|f+g|^{p-1}(|f|+|g|)$ as well as equality in the applications of Hölder's inequality. Equality occurs in in the pointwise inequality above if $f(x)$ and $g(x)$ have the same sign. On the other hand, equality occurs in Hölder's inequality if

$$c_1|f(x)|^p=|f(x)+g(x)|^p=c_2|g(x)|^p \quad \text{for } \mu \text{ almost every } x\in A.$$

This completes the proof. □

Note that the normed space $L^p(A)$, $1\leqslant p<\infty$, is a metric space with the metric

$$d(f,g)=\|f-g\|_p.$$

1.3 L^p spaces for $0 < p < 1$

It is sometimes useful to consider L^p spaces for $0 < p < \infty$. Observe that Definition 1.1 makes sense also when $0 < p < 1$ and the space is a vector space by the same argument as in the proof of Lemma 1.7. However, $\|f\|_p$ is not a norm if $0 < p < 1$.

Reason. Let $f = \chi_{[0,\frac{1}{2})}$ and $g = \chi_{[\frac{1}{2},1]}$. Then $f + g = \chi_{[0,1]}$ so that $\|f + g\|_p = 1$. On the other hand, $\|f\|_p = 2^{-1/p}$ and $\|g\|_p = 2^{-1/p}$. Thus

$$\|f\|_p + \|g\|_p = 2 \cdot 2^{-1/p} = 2^{1-1/p} < 1, \quad \text{when} \quad 0 < p < 1.$$

This shows that $\|f\|_p + \|g\|_p < \|f + g\|_p$. ■

Thus the triangle inequality does not hold true when $0 < p < 1$, but we have the following result.

Lemma 1.24. If $f, g \in L^p(A)$ and $0 < p < 1$, then $f + g \in L^p(A)$ and

$$\|f + g\|_p^p \leqslant \|f\|_p^p + \|g\|_p^p.$$

Proof. The elementary inequality

$$(a + b)^p \leqslant a^p + b^p, \quad a, b \geqslant 0, \quad 0 < p < 1, \tag{1.4}$$

implies

$$\|f + g\|_p^p = \int_A |f + g|^p \, d\mu \leqslant \int_A |f|^p \, d\mu + \int_A |g|^p \, d\mu = \|f\|_p^p + \|g\|_p^p.$$ □

However, $L^p(A)$ is a metric space with the metric

$$d(f, g) = \|f - g\|_p^p = \int_A |f - g|^p \, d\mu$$

This metric is not induced by a norm, since $\|f\|_p^p$ does not satisfy the homogeneity required by the norm. On the other hand, $\|f\|_p$ satisfies the homogeneity, but not the triangle inequality.

Remarks 1.25:

(1) By (1.2), we have

$$\|f + g\|_p \leqslant \left(\|f\|_p^p + \|g\|_p^p\right)^{1/p} \leqslant 2^{1/p}(\|f\|_p + \|g\|_p), \quad 0 < p < 1.$$

Thus a quasi triangle inequality holds with a multiplicative constant.

(2) If $f, g \in L^p(A)$, $f \geqslant 0$, $g \geqslant 0$, then

$$\|f + g\|_p \geqslant \|f\|_p + \|g\|_p, \quad 0 < p < 1.$$

This is the triangle inequality in the wrong direction (exercise).

Remark 1.26. It is possible to define the L^p spaces also when $p<0$. A μ-measurable function is in $L^p(A)$ for $p<0$, if

$$0<\int_A |f|^p\,d\mu=\int_A \frac{1}{|f|^{p'/(1-p')}}\,d\mu<\infty, \quad \text{where} \quad \frac{1}{p}+\frac{1}{p'}=1.$$

If $f\in L^p(A)$ for $p<0$, then $f\neq 0$ μ-almost everywhere and $|f|<\infty$ μ-almost everywhere. However, this is not a vector space.

1.4 Completeness of L^p

Next we shall prove a famous theorem, which is not only important in the integration theory, but has a great historical interest as well. The result was found independently by F. Riesz and E. Fisher in 1907, primarily in connection with the Fourier series which culminates in showing the completeness of L^2.

Recall that a sequence (f_i) of functions $f_i\in L^p(A)$, $i=1,2,\dots$, converges in $L^p(A)$ to a function $f\in L^p(A)$, if for every $\varepsilon>0$ there exists i_ε such that

$$\|f_i-f\|_p<\varepsilon \quad \text{when} \quad i\geqslant i_\varepsilon.$$

Equivalently,

$$\lim_{i\to\infty}\|f_i-f\|_p=0.$$

A sequence (f_i) is a Cauchy sequence in $L^p(A)$, if for every $\varepsilon>0$ there exists i_ε such that

$$\|f_i-f_j\|_p<\varepsilon \quad \text{when} \quad i,j\geqslant i_\varepsilon.$$

WARNING: This is not the same condition as

$$\|f_{i+1}-f_i\|_p<\varepsilon \quad \text{when} \quad i\geqslant i_\varepsilon.$$

Indeed, the Cauchy sequence condition implies this, but the converse is not true (exercise).

CLAIM: If $f_i\to f$ in $L^p(A)$, then (f_i) is Cauchy sequence in $L^p(A)$.

Reason. By Minkowski's inequality

$$\|f_i-f_j\|_p\leqslant\|f_i-f\|_p+\|f-f_j\|_p<\varepsilon$$

when i and j are sufficiently large. ■

Theorem 1.27 (Riesz-Fischer). If (f_i) is a Cauchy sequence in $L^p(A)$, $1\leqslant p<\infty$, then there exists $f\in L^p(A)$ such that $f_i\to f$ in $L^p(A)$ as $i\to\infty$.

THE MORAL: $L^p(A)$, $1\leqslant p<\infty$, is a Banach space with the norm $\|\cdot\|_p$. In particular, $L^2(A)$ is a Hilbert space.

Proof. Assume that (f_i) is a Cauchy sequence in $L^p(A)$. We choose a subsequence as follows. Choose i_1 such that

$$\|f_i - f_j\|_p < \frac{1}{2} \quad \text{when} \quad i, j \geqslant i_1.$$

We continue recursively. Suppose that $i_1, i_2, \dots, i_k$ have been chosen such that

$$\|f_i - f_j\|_p < \frac{1}{2^k} \quad \text{when} \quad i, j \geqslant i_k.$$

Then choose $i_{k+1} > i_k$ such that

$$\|f_i - f_j\|_p < \frac{1}{2^{k+1}} \quad \text{when} \quad i, j \geqslant i_{k+1}.$$

For the subsequence (f_{i_k}), we have

$$\|f_{i_k} - f_{i_{k+1}}\|_p < \frac{1}{2^k}, \quad k = 1, 2, \dots.$$

Define

$$g_l = \sum_{k=1}^{l} |f_{i_{k+1}} - f_{i_k}| \quad \text{and} \quad g = \sum_{k=1}^{\infty} |f_{i_{k+1}} - f_{i_k}|.$$

Then

$$\lim_{l\to\infty} g_l = \lim_{l\to\infty} \sum_{k=1}^{l} |f_{i_{k+1}} - f_{i_k}| = \sum_{k=1}^{\infty} |f_{i_{k+1}} - f_{i_k}| = g$$

and as a limit of μ-measurable functions g is a μ-measurable function. Fatou's lemma and Minkowski's inequality imply

$$\begin{aligned}\left(\int_A g^p \, d\mu\right)^{1/p} &\leqslant \liminf_{l\to\infty} \left(\int_A g_l^p \, d\mu\right)^{1/p} \\ &\leqslant \liminf_{l\to\infty} \sum_{k=1}^{l} \left(\int_A |f_{i_{k+1}} - f_{i_k}|^p \, d\mu\right)^{1/p} \leqslant \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.\end{aligned}$$

Thus $g \in L^p(A)$ and consequently $g(x) < \infty$ for μ-almost every $x \in A$. It follows that the series

$$f_{i_1}(x) + \sum_{k=1}^{\infty} (f_{i_{k+1}}(x) - f_{i_k}(x))$$

converges absolutely for μ-almost every $x \in A$. Denote the sum of the series by $f(x)$ for those $x \in A$ at which it converges and set $f(x) = 0$ in the remaining set of measure zero. Then

$$\begin{aligned}f(x) &= f_{i_1}(x) + \sum_{k=1}^{\infty} (f_{i_{k+1}}(x) - f_{i_k}(x)) \\ &= \lim_{l\to\infty} \left(f_{i_1}(x) + \sum_{k=1}^{l-1} (f_{i_{k+1}}(x) - f_{i_k}(x)) \right) \\ &= \lim_{l\to\infty} f_{i_l}(x) = \lim_{k\to\infty} f_{i_k}(x)\end{aligned}$$

for μ-almost every $x \in A$. Thus there is a subsequence which converges μ-almost everywhere in A. Next we show that the original sequence converges to f in $L^p(A)$.

C L A I M : $f_i \to f$ in $L^p(A)$ as $i \to \infty$.

Reason. Let $\varepsilon > 0$. Since (f_i) is a Caucy sequence in $L^p(A)$, there exists i_ε such that

$$\|f_i - f_j\|_p < \varepsilon \quad \text{when} \quad i,j \geqslant i_\varepsilon.$$

For a fixed i, we have $f_{i_k} - f_i \to f - f_i$ μ-almost everywhere in A as $i_k \to \infty$. By Fatou's lemma

$$\left(\int_A |f - f_i|^p \, d\mu\right)^{1/p} \leqslant \liminf_{k\to\infty} \left(\int_A |f_{i_k} - f_i|^p \, d\mu\right)^{1/p} \leqslant \varepsilon.$$

■

This shows that $f - f_i \in L^p(A)$ and thus $f = (f - f_i) + f_i \in L^p(A)$. Moreover, for every $\varepsilon > 0$ there exists i_ε such that

$$\|f_i - f\|_p < \varepsilon \quad \text{when} \quad i \geqslant i_\varepsilon.$$

This completes the proof. □

WARNING: In general, if a sequence has a converging subsequence, the original sequence need not converge. In the proof above, we used the fact that we have a Cauchy sequence.

We shall often use a particular part of the proof of the Riesz-Fisher theorem, which we now state.

Corollary 1.28. If $f_i \to f$ in $L^p(A)$, then there exist a subsequence (f_{i_k}) such that

$$\lim_{k\to\infty} f_{i_k}(x) = f(x) \quad \mu\text{-almost every} \quad x \in A.$$

Proof. The proof of the Riesz-Fischer theorem gives a subsequence (f_{i_k}) and a function $g \in L^p(A)$ such that

$$\lim_{k\to\infty} f_{i_k}(x) = g(x) \quad \mu\text{-almost every} \quad x \in A$$

and $f_{i_k} \to g$ in $L^p(A)$. On the other hand, $f_i \to f$ in $L^p(A)$, which implies that $f_{i_k} \to f$ in $L^p(A)$. By the uniqueness of the limit, we conclude that $f = g$ μ-almost everywhere in A. □

Remarks 1.29:

Let us compare the various modes of convergence of a sequence (f_i) of functions in $L^p(A)$.

(1) If $f_i \to f$ in $L^p(A)$, then

$$\lim_{i\to\infty} \|f_i\|_p = \|f\|_p.$$

Reason. $\|f_i\|_p = \|f_i - f + f\|_p \leqslant \|f_i - f\|_p + \|f\|_p$ implies $\|f_i\|_p - \|f\|_p \leqslant \|f_i - f\|_p$. In the same way $\|f\|_p - \|f_i\|_p \leqslant \|f_i - f\|_p$. Thus

$$|\|f_i\|_p - \|f\|_p| \leqslant \|f_i - f\|_p \to 0,$$

from which it follows that

$$\lim_{i\to\infty} \|f_i\|_p = \|f\|_p.$$

■

(2) $f_i \to f$ in $L^p(A)$ implies that $f \to f$ in measure.

Reason. By Chebyshev's inequality

$$\begin{aligned}\mu(\{x \in A : |f_i(x) - f(x)| \geqslant \varepsilon\}) &\leqslant \frac{1}{\varepsilon^p}\int_A |f_i - f|^p \, d\mu \\ &= \frac{1}{\varepsilon^p}\|f_i - f\|_p \to 0 \quad \text{as} \quad i \to \infty.\end{aligned}$$ ■

(3) If $f_i \to f$ in $L^p(A)$, then there exist a subsequence (f_{i_k}) such that

$$\lim_{k\to\infty} f_{i_k}(x) = f(x) \quad \mu\text{-almost every} \quad x \in A.$$

Reason. The convergence in measure implies the existence of an almost everywhere converging subsequence. This gives another proof of the previous corollary ■

(4) In the case $p = 1$, $f_i \to f$ in $L^1(A)$ implies not only that

$$\lim_{i\to\infty}\int_A |f_i| \, d\mu = \int_A |f| \, d\mu$$

but also that

$$\lim_{i\to\infty}\int_A f_i, d\mu = \int_A f \, d\mu.$$

Reason.

$$\left|\int_A (f_i - f)\, d\mu\right| \leqslant \int_A |f_i - f| \, d\mu = \|f_i - f\|_1 \to 0 \quad \text{as} \quad i \to \infty.$$ ■

THE MORAL: This is a useful tool in showing that a sequence does not converge in L^p.

Example 1.30. Let $f_i = \chi_{[i-1,i)}$, $i = 1, 2, \ldots$, and $f = 0$. Assume that μ is the Lebesgue measure. Then

$$\lim_{i\to\infty} f_i(x) = f(x) \quad \text{for every} \quad x \in \mathbb{R}.$$

However, $\|f_i\|_p = 1$ for every $i = 1, 2, \ldots$ and $\|f\|_p = 0$. Thus the sequence (f_i) does not converge to f in $L^p(\mathbb{R})$, $1 \leqslant p < \infty$.

Example 1.31. In the following examples we assume that μ is the Lebesgue measure.

(1) $f_i \to f$ almost everywhere does not imply $f_i \to f$ in L^p. Let

$$f_i = i^2 \chi_{(0,\frac{1}{i})}, \quad i = 1, 2, \ldots.$$

Then

$$\int_{\mathbb{R}} |f_i|^p \, dx = i^{2p}\int_{\mathbb{R}} \chi^p_{(0,\frac{1}{i})} \, dx = i^{2p}\frac{1}{i} = i^{2p-1} < \infty.$$

Thus $f_i \in L^p(\mathbb{R})$, $1 \leqslant p < \infty$, $f_i(x) \to 0$ for every $x \in \mathbb{R}$, but

$$\|f_i\|_p = i^{2-\frac{1}{p}} \geqslant i \to \infty \quad \text{as} \quad i \to \infty.$$

Thus (f_i) does not converge in $L^p(\mathbb{R})$.

(2) $f_i \to f$ in L^p does not imply $f_i \to f$ almost everywhere. Consider the sliding sequence of functions

$$f_{2^k+j} = k\chi_{\left[\frac{j}{2^k},\frac{j+1}{2^k}\right]}, \quad k = 0,1,2,\dots, \quad j = 0,1,2,\dots,2^k - 1.$$

Then

$$\|f_{2^k+j}\|_p = k2^{-\frac{k}{p}} \to 0 \quad \text{as} \quad k \to \infty$$

which implies that $f_i \to 0$ in $L^p(\mathbb{R})$, $1 \leqslant p < \infty$, as $i \to \infty$. However, the sequence $(f_i(x))$ fails to converge for every $x \in [0,1]$, since

$$\limsup_{i\to\infty} f_i(x) = \infty \quad \text{and} \quad \liminf_{i\to\infty} f_i(x) = 0$$

for every $x \in [0,1]$. Note that there are many converging subsequences. For example, $f_{2^k+1}(x) \to 0$ for every $x \in [0,1]$ as $k \to \infty$.

(3) A sequence can converge in L^p without converging in L^q. Consider

$$f_i = i^{-1}\chi_{(i,2i)}, \quad i = 1,2,\dots.$$

Then $\|f_i\|_p = i^{-1+1/p}$. Thus $f \to 0$ in $L^p(\mathbb{R})$, $1 < p < \infty$, but $\|f_i\|_1 = 1$ for every $i = 1,2\dots$, so that the sequence (f_i) does not converge in $L^1(\mathbb{R}^n)$.

The following theorem clarifies the difference between the pointwise convengence and L^p-convergence.

Theorem 1.32. Assume that $f_i \in L^p(A)$, $i = 1,2,\dots$ and $f \in L^p(A)$, $1 \leqslant p < \infty$. If $f_i \to f$ μ-almost everywhere in A and $\lim_{i\to\infty} \|f_i\|_p = \|f\|_p$, then $f_i \to f$ in $L^p(A)$ as $i \to \infty$.

Proof. Since $|f_i| < \infty$ and $|f| < \infty$ μ-almost everywhere in A, by (1.2), we have

$$2^p(|f_i|^p + |f|^p) - |f_i - f|^p \geqslant 0 \quad \mu\text{-almost everywhere in } A.$$

The assumption $f_i \to f$ μ-almost everywhere in A implies

$$\lim_{i\to\infty}(2^p(|f_i|^p + |f|^p) - |f_i - f|^p) = 2^{p+1}|f|^p \quad \mu\text{-almost everywhere in } A.$$

Applying Fatou's lemma, we obtain

$$\begin{aligned}
\int_A 2^{p+1}|f|\,d\mu &\leqslant \liminf_{i\to\infty}\int_A \left(2^p(|f_i|^p + |f|^p) - |f_i - f|^p\right) d\mu \\
&\leqslant \liminf_{i\to\infty}\left(\int_A 2^p|f_i|^p\,d\mu + \int_A 2^p|f|^p\,d\mu - \int_A |f_i - f|^p\,d\mu\right) \\
&= \lim_{i\to\infty}\int_A 2^p|f_i|^p\,d\mu + \int_A 2^p|f|^p\,d\mu - \limsup_{i\to\infty}\int_A |f_i - f|^p\,d\mu \\
&= \int_A 2^p|f|^p\,d\mu + \int_A 2^p|f|^p\,d\mu - \limsup_{i\to\infty}\int_A |f_i - f|^p\,d\mu.
\end{aligned}$$

Here we used the facts that if (a_i) is a converging sequence of real numbers and (b_i) is an arbitrary sequence of real numbers, then

$$\liminf_{i\to\infty}(a_i+b_i)=\lim_{i\to\infty}a_i+\liminf_{i\to\infty}b_i \quad \text{and} \quad \liminf_{i\to\infty}(-b_i)=-\limsup_{i\to\infty}b_i.$$

Subtracting $\int_A 2^{p+1}|f|^p\,d\mu$ from both sides, we have

$$\limsup_{i\to\infty}\int_A |f_i-f|^p\,d\mu\leqslant 0.$$

On the other hand, since the integrands are nonnegative

$$\limsup_{i\to\infty}\int_A |f_i-f|^p\,d\mu\leqslant 0.$$

Thus

$$\lim_{i\to\infty}\int_A |f_i-f|^p\,d\mu=0.$$

□

1.5 L^∞ space

The definition of the L^∞ space differs substantially from the definition of the L^p space for $1\leqslant p<\infty$. The main difference is that instead of the integration the definition is based on the almost everywhere concept. The class L^∞ consists of bounded measurable functions with the interpretation that we neglect the behaviour of the functions on a set of measure zero.

Definition 1.33. Let $A\subset\mathbb{R}^n$ be a μ-measurable set and $f:A\to[-\infty,\infty]$ a μ-measurable function. Then $f\in L^\infty(A)$, if there exists M, $0\leqslant M<\infty$, such that

$$|f(x)|\leqslant M \quad \text{for } \mu\text{-almost every } x\in A.$$

Functions in L^∞ are sometimes called essentially bounded functions. If $f\in L^\infty(A)$, then the essential supremum of f is

$$\begin{aligned}\operatorname*{ess\,sup}_{x\in A}|f(x)| &= \inf\{M : |f(x)|\leqslant M \text{ for } \mu\text{-almost every } x\in A\}\\ &= \inf\left\{M : \mu(\{x\in A : |f(x)|>M\})=0\right\}\end{aligned}$$

and the essential infimum of f is

$$\begin{aligned}\operatorname*{ess\,inf}_{x\in A}|f(x)| &= \sup\{m : |f(x)|\geqslant m \text{ for } \mu\text{-almost every } x\in A\}\\ &= \sup\left\{m : \mu(\{x\in A : |f(x)|<m\})=0\right\}.\end{aligned}$$

The L^∞ norm of f is

$$\|f\|_\infty=\operatorname*{ess\,sup}_{x\in A}|f(x)|.$$

It is clear that $f\in L^\infty(A)$ if and only if $\|f\|_\infty<\infty$.

THE MORAL: $\|f\|_\infty$ is the supremum outside sets of measure zero. Observe that the standard supremum of a bounded function f is

$$\sup_{x\in A}|f(x)| = \inf\{M : \{x \in A : |f(x)| > M\} = \emptyset\}.$$

WARNING: The L^p norm for $1 \leqslant p < \infty$ depends on the average size of the function, but L^∞ norm depends on the pointwise values of the function outside a set of measure zero. More precisely, the L^p norm for $1 \leqslant p < \infty$ depends very much on the underlying measure μ and would be very sensitive to any changes in μ. The L^∞ depends only on the class of sets of μ measure zero and not on the distribution of the measure μ itself.

Remark 1.34. In the special case that $A = \mathbb{N}$ and μ is the counting measure, the $L^\infty(\mathbb{N})$ space is denoted by l^∞ and

$$l^\infty = \left\{(x_i) : \sup_{i\in\mathbb{N}}|x_i| < \infty\right\}.$$

Here (x_i) is a sequence of real (or complex) numbers. Thus l^∞ is the space of bounded sequences.

Example 1.35. Assume that μ is the Lebesgue measure.

(1) Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \chi_{\mathbb{Q}}(x)$. Then $\|f\|_\infty = 0$, but $\sup_{x\in\mathbb{R}}|f(x)| = 1$.
(2) Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = 1/|x|$. Then $f \notin L^\infty(\mathbb{R}^n)$.

Remarks 1.36:
(1) $\|f\|_\infty \leqslant \sup_{x\in A}|f(x)|$.
(2) Let $f \in L^\infty(A)$. Then for every $\varepsilon > 0$, we have

$$\mu(\{x \in A : |f(x)| > \|f\|_\infty + \varepsilon\}) = 0 \quad \text{and} \quad \mu(\{x \in A : |f(x)| > \|f\|_\infty - \varepsilon\}) > 0.$$

(3) If $f \in C(A)$ and $\mu(A) > 0$, then $\|f\|_\infty = \sup_{x\in A}|f(x)|$. (Exercise)

Lemma 1.37. Assume that $f \in L^\infty(A)$. Then

(1) $|f(x)| \leqslant \operatorname{ess\,sup}_{x\in A}|f(x)|$ for μ-almost every $x \in A$ and
(2) $|f(x)| \geqslant \operatorname{ess\,inf}_{x\in A}|f(x)|$ for μ-almost every $x \in A$.

THE MORAL: In other words, $|f(x)| \leqslant \|f\|_\infty$ for μ-almost every $x \in A$. This means that if $f \in L^\infty$, there exists a smallest number number M such that $|f(x)| \leqslant M$ for μ-almost every $x \in A$. This smallest number is $\|f\|_\infty$.

Proof. (1) For every $i = 1, 2, \dots$ there exists $M_i \geqslant 0$ such that

$$M_i < \|f\|_\infty + \frac{1}{i} \quad \text{and} \quad |f(x)| \leqslant M_i \quad \text{for } \mu\text{-almost every } x \in A.$$

Thus there exists $N_i \subset A$ with $\mu(N_i) = 0$ such that $|f(x)| \leqslant M_i$ for every $x \in A \setminus N_i$. Let $N = \bigcup_{i=1}^{\infty} N_i$. Then $\mu(N) \leqslant \sum_{i=1}^{\infty} \mu(N_i) = 0$. Observe that

$$\bigcap_{i=1}^{\infty}(A \setminus N_i) = A \setminus \bigcup_{i=1}^{\infty} N_i = A \setminus N.$$

Then

$$|f(x)| \leqslant M_i < \|f\|_\infty + \frac{1}{i} \quad \text{for every} \quad x \in A \setminus N, \quad i = 1, 2, \dots.$$

Letting $i \to \infty$, we obtain $|f(x)| \leqslant \|f\|_\infty$ for every $x \in A \setminus N$.

(2) (Exercise) □

Lemma 1.38 (Minkowski's inequality for $p = \infty$). If $f, g \in L^\infty(A)$, then

$$\|f + g\|_\infty \leqslant \|f\|_\infty + \|g\|_\infty.$$

Proof. By Lemma 1.37 $|f(x)| \leqslant \|f\|_\infty$ for μ-almost every $x \in A$ and $|g(x)| \leqslant \|g\|_\infty$ for μ-almost every $x \in A$. Thus

$$|f(x) + g(x)| \leqslant |f(x)| + |g(x)| \leqslant \|f\|_\infty + \|g\|_\infty \quad \text{for } \mu\text{-almost every } x \in A.$$

By the definition of the L^∞ norm, we have $\|f + g\|_\infty \leqslant \|f\|_\infty + \|g\|_\infty$. □

THE MORAL: This is the triangle inequality for the L^∞-norm. It implies that the L^∞ norm is a norm in the usual sense and that $L^\infty(A)$ is a normed space if the functions that coincide almost everywhere are identified.

Theorem 1.39 (Hölder's inequality for $p = \infty$ and $p' = 1$). If $f \in L^1(A)$ ja $g \in L^\infty(A)$, then $fg \in L^1(A)$

$$\|fg\|_1 \leqslant \|g\|_\infty \|f\|_1.$$

THE MORAL: In practice, we take the essential supremum out of the integral.

Proof. By Lemma 1.37, we have $|g(x)| \leqslant \|g\|_\infty$ for μ-almost every $x \in A$. This implies

$$|f(x)g(x)| \leqslant \|g\|_\infty |f(x)| \quad \text{for } \mu\text{-almost every } x \in A$$

and thus

$$\int_A |f(x)g(x)|\, d\mu \leqslant \|g\|_\infty \|f\|_1.$$

□

Remark 1.40. There is also an L^p version $\|fg\|_p \leqslant \|g\|_\infty \|f\|_p$ of the previous theorem.

Next result justifies the notation $\|f\|_\infty$.

Theorem 1.41. If $f \in L^p(A)$ for some $1 \leqslant p < \infty$, then

$$\lim_{p \to \infty} \|f\|_p = \|f\|_\infty.$$

THE MORAL: In this sense, $L^\infty(A)$ is the limit of $L^p(A)$ spaces as $p \to \infty$. Moreover, this gives a useful method to show that $f \in L^\infty$: It is enough find a uniform bound for the L^p norms as $p \to \infty$.

Proof. Denote $A_\lambda = \{x \in A : |f(x)| > \lambda\}$, $\lambda \geqslant 0$. Suppose $0 \leqslant \lambda < \|f\|_\infty$. By the definition of the L^∞ norm, we have $\mu(A_\lambda) > 0$. By Chebyshev's inequality

$$\mu(A_\lambda) \leqslant \int_A \left(\frac{|f|}{\lambda}\right)^p d\mu = \frac{1}{\lambda^p}\int_A |f|^p\, d\mu < \infty$$

and thus $\|f\|_p \geqslant \lambda\mu(A_\lambda)^{1/p}$. Since $0 < \mu(A_\lambda) < \infty$, we have $\mu(A_\lambda)^{1/p} \to 1$ as $p \to \infty$. This implies

$$\liminf_{p\to\infty} \|f\|_p \geqslant \lambda \quad \text{whenever} \quad 0 \leqslant \lambda < \|f\|_\infty.$$

By letting $\lambda \to \|f\|_\infty$, we have

$$\liminf_{p\to\infty} \|f\|_p \geqslant \|f\|_\infty.$$

On the other hand, for $q < p < \infty$, we have

$$\|f\|_p = \left(\int_A |f|^p\, d\mu\right)^{1/p} = \left(\int_A |f|^q |f|^{p-q}\, d\mu\right)^{1/p} \leqslant \|f\|_\infty^{1-q/p} \|f\|_q^{q/p}.$$

Since $\|f\|_q < \infty$ for some q, this implies

$$\limsup_{p\to\infty} \|f\|_p \leqslant \|f\|_\infty.$$

We have shown that

$$\limsup_{p\to\infty} \|f\|_p \leqslant \|f\|_\infty \leqslant \liminf_{p\to\infty} \|f\|_p,$$

which implies that the limit exists and

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty.$$

□

Remarks 1.42:

(1) The assumption $f \in L^p(A)$ for some $1 \leqslant p < \infty$ can be replaced with the assumption $\mu(A) < \infty$.

(2) Recall that by Jensen's inequality, the integral average

$$\left(\frac{1}{\mu(A)}\int_A |f|^p\, d\mu\right)^{1/p}$$

is an increasing function of p.

(3) If $0 < \mu(A) < \infty$, then for every μ-measurable function

$$\lim_{p\to\infty}\left(\frac{1}{\mu(A)}\int_A |f|^p\, d\mu\right)^{1/p} = \operatorname*{ess\,sup}_A |f|,$$

$$\lim_{p\to\infty}\left(\frac{1}{\mu(A)}\int_A |f|^{-p}\,d\mu\right)^{-1/p} = \operatorname*{ess\,inf}_A |f|$$

and

$$\lim_{p\to 0}\left(\frac{1}{\mu(A)}\int_A |f|^{p}\,d\mu\right)^{1/p} = \exp\left(\frac{1}{\mu(A)}\int_A \log|f|\,d\mu\right).$$

Theorem 1.43. $L^\infty(A)$ is a Banach space.

T H E M O R A L : The claim and proof is the same as in showing that the space of continuous functions with the supremum norm is complete. The only difference is that we have to neglect sets of zero measure.

Proof. Let (f_i) be a Cauchy sequence in $L^\infty(A)$. By Lemma 1.37, we have

$$|f_i(x) - f_j(x)| \leqslant \|f_i - f_j\|_\infty \quad \text{for } \mu\text{-almost every } x \in A.$$

Thus there exists $N_{i,j} \subset A$, $\mu(N_{i,j}) = 0$ such that

$$|f_i(x) - f_j(x)| \leqslant \|f_i - f_j\|_\infty \quad \text{for every } x \in A \setminus N_{i,j}.$$

Since (f_i) is a Cauchy sequence in $L^\infty(A)$, for every $k = 1, 2, \ldots$, there exists i_k such that

$$\|f_i - f_j\|_\infty < \frac{1}{k} \quad \text{when} \quad i, j \geqslant i_k.$$

This implies

$$|f_i(x) - f_j(x)| < \frac{1}{k} \quad \text{for every } x \in A \setminus N_{i,j}, \quad i, j \geqslant i_k.$$

Let $N = \bigcup_{i=1}^\infty \bigcup_{j=1}^\infty N_{i,j}$. Then

$$\mu(N) \leqslant \sum_{i=1}^\infty \sum_{j=1}^\infty \mu(N_{i,j}) = 0$$

and

$$|f_i(x) - f_j(x)| < \frac{1}{k} \quad \text{for every } x \in A \setminus N, \quad i, j \geqslant i_k.$$

Thus $(f_i(x))$ is a Cauchy sequence for every $x \in A \setminus N$. Since $\mathbb{R}$ is complete, there exists

$$\lim_{i\to\infty} f_i(x) = f(x) \quad \text{for every} \quad x \in A \setminus N.$$

We set $f(x) = 0$, when $x \in N$. Then f is measurable as a pointwise limit of measurable functions Letting $j \to \infty$ in the preceding inequality gives

$$|f_i(x) - f(x)| \leqslant \frac{1}{k} \quad \text{for every } x \in A \setminus N_{i,j}, \quad i, j \geqslant i_k,$$

which implies

$$\|f_i - f\|_\infty \leqslant \frac{1}{k} \quad \text{when} \quad i \geqslant i_k.$$

Since $\|f\|_\infty \leqslant \|f_i\|_\infty + \|f_i - f\|_\infty < \infty$, we have $f \in L^\infty(A)$ and $f_i \to f$ in $L^\infty(A)$ as $i \to \infty$. □

Remark 1.44. The proof shows that $f_i \to f$ in $L^\infty(A)$ as $i \to \infty$ implies that $f_i \to f$ uniformly in $A \setminus N$ with $\mu(N) = 0$.

Example 1.45. Assume that μ is the Lebesgue measure. Let $f_i : \mathbb{R} \to \mathbb{R}$,

$$f_i(x) = \begin{cases} 0, & x \in (-\infty, 0), \\ ix, & x \in \left[0, \frac{1}{i}\right], \\ 1, & x \in \left(\frac{1}{i}, \infty\right), \end{cases}$$

for $i = 1, 2, \dots$ and let $f = \chi_{[0,\infty)}$. Then $f_i(x) \to f(x)$ for every $x \in \mathbb{R}$ as $i \to \infty$, $\|f_i\|_\infty = 1$ for every $i = 1, 2, \dots$, $\|f\|_\infty = 1$ so that $\lim_{i\to\infty} \|f_i\|_\infty = \|f\|_\infty$, but $\|f_i - f\|_\infty = 1$ for every $i = 1, 2, \dots$. Thus $\lim_{i\to\infty} \|f_i - f\|_\infty = 1 \neq 0$. This shows that the claim of Theorem 1.32 does not hold when $p = \infty$.

The Hardy-Littlewood maximal function is a very useful tool in analysis. The maximal function theorem asserts that the maximal operator is bounded from L^p to L^p for $p > 1$ and for $p = 1$ there is a weak type estimate. The weak type estimate is used to prove the Lebesgue differentiation theorem, which gives a pointwise meaning for a locally integrable function. The Lebesgue differentiation theorem is a higher dimensional version of the fundamental theorem of calculus. It is applied to the study of the density points of a measurable set. As an application we prove a Sobolev embedding theorem.

2 The Hardy-Littlewood maximal function

In this section we restrict our attention to the Lebesgue measure on $\mathbb{R}^n$. We prove Lebesgue's theorem on differentiation of integrals, which is an extension of the one-dimensional fundamental theorem of calculus to the n-dimensional case. This theorem states that, for a (locally) integrable function $f : \mathbb{R}^n \to [-\infty, \infty]$, we have

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

for almost every $x \in \mathbb{R}^n$. Recall that $B(x,r) = \{y \in \mathbb{R}^n : |y - x| < r\}$ is the open ball with the center x and radius $r > 0$. In proving this result we need to investigate very carefully the behaviour of the integral averages above. This leads to the Hardy-Littlewood maximal function, where we take the supremum of the integral averages instead of the limit. The passage from the limiting expression to a corresponding maximal function is a situation that occurs often. Hardy and Littlewood wrote that they were led to study the one-dimensional version of the maximal function by the question how a score in cricket can be maximized: "*The problem is most easily grasped when stated in the language of cricket, or any other game in which the player complies a series of scores of which average is recorded.*"As we shall see, these concepts and methods have a universal significance in analysis.

2.1 Local L^p spaces

If we are interested in pointwise properties of functions, it is not necessary to require integrablity conditions over the whole underlying domain. For example, in the Lebesgue differentiation theorem above, the limit is taken over integral averages over balls that shrink to the point x, so that the behaviour of f far away from x is irrelevant.

Definition 2.1. Let $\Omega \subset \mathbb{R}^n$ be an open set and assume that $f : \Omega \to [-\infty, \infty]$ is a measurable function. Then $f \in L^p_{\text{loc}}(\Omega)$, if

$$\int_K |f|^p \, dx < \infty, \quad 1 \leqslant p < \infty,$$

and

$$\operatorname*{ess\,sup}_K |f| < \infty, \quad p = \infty$$

for every compact set $K \subset \Omega$.

Examples 2.2:

$L^p(\Omega) \subset L^p_{\text{loc}}(\Omega)$, but the reverse inclusion is not true.

(1) Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = 1$. Then $f \notin L^p(\mathbb{R}^n)$ for any $1 \leqslant p < \infty$, but $f \in L^p_{\text{loc}}(\mathbb{R}^n)$ for every $1 \leqslant p < \infty$.

(2) Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = |x|^{-1/2}$. Then $f \notin L^1(\mathbb{R}^n)$, but $f \in L^1_{\text{loc}}(\mathbb{R}^n)$.

(3) Let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = e^{|x|}$. Then $f \notin L^1(\mathbb{R}^n)$, but $f \in L^1_{\text{loc}}(\mathbb{R}^n)$.

(4) Let $f : B(0,1) \setminus \{0\} \to \mathbb{R}$, $f(x) = |x|^{-n/p}$. Then $f \notin L^p(B(0,1) \setminus \{0\})$ for $1 \leqslant p < n$, but $f \in L^p_{\text{loc}}(B(0,1) \setminus \{0\})$ for $1 < p < \infty$. Moreover, $f \notin L^\infty(B(0,1) \setminus \{0\})$, but $f \in L^\infty_{\text{loc}}(B(0,1) \setminus \{0\})$.

(5) For $p = \infty$, let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = |x|$. Then $f \notin L^\infty(\mathbb{R}^n)$, but $f \in L^\infty_{\text{loc}}(\mathbb{R}^n)$.

Remarks 2.3:

(1) If $1 \leqslant p \leqslant q \leqslant \infty$, then $L^\infty_{\text{loc}}(\Omega) \subset L^q_{\text{loc}}(\Omega) \subset L^p_{\text{loc}}(\Omega) \subset L^1_{\text{loc}}(\Omega)$.

Reason. By Jensen's inequality

$$\frac{1}{|K|} \int_K |f| \, dx \leqslant \left(\frac{1}{|K|} \int_K |f|^p \, dx \right)^{1/p} \leqslant \left(\frac{1}{|K|} \int_K |f|^q \, dx \right)^{1/q} \leqslant \operatorname*{ess\,sup}_K |f|,$$

where K is a compact subset of Ω with $|K| > 0$. ∎

(2) $C(\Omega) \subset L^p_{\text{loc}}(\Omega)$ for every $1 \leqslant p \leqslant \infty$.

Reason. Since $|f| \in C(\Omega)$ assumes its maximum in the compact set K and K has a finite Lebesgue measure, we have

$$\int_K |f|^p \, dx \leqslant |K| (\operatorname*{ess\,sup}_K |f|)^p \leqslant |K| (\max_K |f|)^p < \infty.$$

∎

(3) $f \in L^p_{\text{loc}}(\mathbb{R}^n) \iff f \in L^p(B(0,r))$ for every $0 < r < \infty \iff f \in L^p(A)$ for every bounded measurable set $A \subset \mathbb{R}^n$.

(4) In general, the quantity

$$\sup_{K \subset \mathbb{R}^n} \left(\int_K |f|^p \, dx \right)^{1/p}$$

is not a norm in $L^p_{\text{loc}}(\mathbb{R}^n)$, since it may be infinity for some $f \in L^p_{\text{loc}}(\mathbb{R}^n)$. Consider, for example, constant functions on $\mathbb{R}^n$.

2.2 Definition of the maximal function

We begin with the definition of the maximal function.

Definition 2.4. The centered Hardy-Littlewood maximal function $Mf : \mathbb{R}^n \to [0,\infty]$ of $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ is defined by

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

where $B(x,r) = \{y \in \mathbb{R}^n : |y-x| < r\}$ is the open ball with the radius $r > 0$ and the center $x \in \mathbb{R}^n$.

T H E M O R A L : $Mf(x)$ gives the maximal integral average of the absolute value of the function on balls centered at x.

Remarks 2.5:

(1) It is enough to assume that $f : \mathbb{R}^n \to [-\infty,\infty]$ is a measurable function in the definition of the Hardy-Littlewood maximal function. The assumption $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ guarantees that the integral averages are finite.

(2) Mf is defined at every point $x \in \mathbb{R}^n$. If $f = g$ almost everywhere in $\mathbb{R}^n$, then $Mf(x) = Mg(x)$ for every $x \in \mathbb{R}^n$.

(3) It may happen that $Mf(x) = \infty$ for every $x \in \mathbb{R}^n$. For example, let $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = |x|$. Then $Mf(x) = \infty$ for every $x \in \mathbb{R}^n$.

(4) There are several seemingly different definitions, which are comparable. Let

$$\widetilde{M}f(x) = \sup_{B \ni x} \frac{1}{|B|} \int_B |f(y)|\,dy$$

be the noncentered maximal function, where the supremum is taken over all open balls B containing the point $x \in \mathbb{R}^n$, then

$$Mf(x) \leqslant \widetilde{M}f(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

On the other hand, if $B = B(z,r) \ni x$, then $B(z,r) \subset B(x,2r)$ and

$$\begin{aligned} \frac{1}{|B|} \int_B |f(y)|\,dy &\leqslant \frac{|B(x,2r)|}{|B(z,r)|} \frac{1}{|B(x,2r)|} \int_{B(x,2r)} |f(y)|\,dy \\ &= 2^n \frac{1}{|B(x,2r)|} \int_{B(x,2r)} |f(y)|\,dy \\ &\leqslant 2^n Mf(x). \end{aligned}$$

This implies that $\widetilde{M}f(x) \leqslant 2^n Mf(x)$ and thus

$$Mf(x) \leqslant \widetilde{M}f(x) \leqslant 2^n Mf(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

(5) It is possible to use cubes in the definition of the maximal function and this will give a comparable notion as well.

Examples 2.6:

(1) Let $f:\mathbb{R}\to\mathbb{R}$, $f=\chi_{[a,b]}$. Then $Mf(x)=1$, if $x\in(a,b)$. For $x\geqslant b$ a calculation shows that the maximal average is obtained when $r=x-a$. Similarly, when $x\leqslant a$, the maximal average is obtained when $r=b-x$. Thus

$$Mf(x)=\begin{cases}\frac{b-a}{2|x-b|}, & x\leqslant a,\\ 1, & x\in(a,b),\\ \frac{b-a}{2|x-a|}, & x\geqslant b.\end{cases}$$

Note that the centered maximal function Mf has jump discontinuities at $x=a$ and $x=b$.

THE MORAL: $f\in L^1(\mathbb{R})$ does not imply $Mf\in L^1(\mathbb{R})$.

(2) Consider the noncenter maximal function $\widetilde{M}f$ of $f:\mathbb{R}\to\mathbb{R}$, $f=\chi_{[a,b]}$. Again $\widetilde{M}f(x)=1$, if $x\in(a,b)$. For $x>b$ a calculation shows that the maximal average over all intervals $(z-r,z+r)$ is obtained when $z=(x+a)/2$ and $r=(x-a)/2$. Similarly, when $x<a$, the maximal average is obtained when $z=(b+x)/2$ and $r=(b-x)/2$. Thus

$$\widetilde{M}f(x)=\begin{cases}\frac{b-a}{|x-b|}, & x\leqslant a,\\ 1, & x\in(a,b),\\ \frac{b-a}{|x-a|}, & x\geqslant b.\end{cases}$$

Note that the uncentered maximal function Mf does not have discontinuities at $x=a$ and $x=b$.

Lemma 2.7. If $f\in C(\mathbb{R}^n)$, then $|f(x)|\leqslant Mf(x)$ for every $x\in\mathbb{R}^n$.

THE MORAL: This justifies the terminology, since the maximal function is pointwise bigger or equal than the absolute value of the original function.

Proof. Assume that $f\in C(\mathbb{R}^n)$ and let $x\in\mathbb{R}^n$. Then for every $\varepsilon>0$ there exists $\delta>0$ such that $|f(x)-f(y)|<\varepsilon$ if $|x-y|<\delta$. This implies

$$\begin{aligned}\left|\frac{1}{|B(x,r)|}\int_{B(x,r)}|f(y)|\,dy-|f(x)|\right| &= \left|\frac{1}{|B(x,r)|}\int_{B(x,r)}(|f(y)|-|f(x)|)\,dy\right|\\ &\leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)}\big||f(y)|-|f(x)|\big|\,dy\\ &\leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)}|f(y)-f(x)|\,dy\leqslant\varepsilon,\quad\text{if}\quad r\leqslant\delta.\end{aligned}$$

Thus

$$|f(x)|=\lim_{r\to0}\frac{1}{|B(x,r)|}\int_{B(x,r)}|f(y)|\,dy\leqslant Mf(x)\quad\text{for every}\quad x\in\mathbb{R}^n.$$

□

The next thing we would like to show is that $Mf:\mathbb{R}^n\to[0,\infty]$ is a measurable function. Recall that a function $f:\mathbb{R}^n\to[-\infty,\infty]$ is lower semicontinuous, if the distribution set $\{x \in \mathbb{R}^n : f(x) > \lambda\}$ is open for every $\lambda \in \mathbb{R}$. Since open sets are Lebesgue measurable, it follows that every lower semicontinuous function is Lebesgue measurable.

Lemma 2.8. Mf is lower semicontinuous.

Proof. Let $A_\lambda = \{x \in \mathbb{R}^n : Mf(x) > \lambda\}$, $\lambda > 0$. For every $x \in A_\lambda$ there exists $r > 0$ such that

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy > \lambda.$$

By the properties of the integral

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy = \lim_{\substack{r'\to r\\ r'>r}} \frac{1}{|B(x,r')|}\int_{B(x,r)} |f(y)|\,dy,$$

which implies that there exists $r' > r$ such that

$$\frac{1}{|B(x,r')|}\int_{B(x,r)} |f(y)|\,dy > \lambda.$$

If $|x-x'| < r'-r$, then $B(x,r) \subset B(x',r')$, since $|y-x'| \leqslant |y-x|+|x-x'| < r+(r'-r) = r'$ for every $y \in B(x,r)$. Thus

$$\begin{aligned}\lambda &< \frac{1}{|B(x,r')|}\int_{B(x,r)} |f(y)|\,dy \leqslant \frac{1}{|B(x,r')|}\int_{B(x',r')} |f(y)|\,dy\\ &= \frac{1}{|B(x',r')|}\int_{B(x',r')} |f(y)|\,dy \leqslant Mf(x'), \quad \text{if} \quad |x-x'| < r'-r.\end{aligned}$$

This shows that $B(x,r'-r) \subset A_\lambda$ and thus A_λ is an open set. □

2.3 Hardy-Littlewood-Wiener maximal function theorems

Another point of view is to consider the Hardy-Littlewood maximal operator $f \mapsto Mf$. We shall list some properties of this operator below.

Lemma 2.9. Assume that $f, g \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

(1) (Positivity) $Mf(x) \geqslant 0$ for every $x \in \mathbb{R}^n$.
(2) (Sublinearity) $M(f+g)(x) \leqslant Mf(x) + Mg(x)$.
(3) (Homogeneity) $M(af)(x) = |a|Mf(x)$, $a \in \mathbb{R}$.
(4) (Translation invariance) $M(\tau_y f)(x) = (\tau_y Mf)(x)$, $y \in \mathbb{R}^n$, where $\tau_y f(x) = f(x+y)$.

Proof. Exercise. □

We are interested in behaviour of the maximal operator in L^p-spaces. The following results were first proved by Hardy and Littlewood in the one-dimensional case and extended later by Wiener to the higher dimensional case.

Lemma 2.10. If $f \in L^\infty(\mathbb{R}^n)$, then $Mf \in L^\infty(\mathbb{R}^n)$ and $\|Mf\|_\infty \leqslant \|f\|_\infty$.

THE MORAL: The maximal function is essentially bounded, and thus finite almost everywhere, if the original function is essentially bounded. Intuitively this is clear, since the integral averages cannot be bigger than the essential supremum of the function.

Proof. For every $x \in \mathbb{R}^n$ and $r > 0$ we have

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy \leqslant \frac{1}{|B(x,r)|}\|f\|_\infty |B(x,r)| = \|f\|_\infty.$$

Thus $Mf(x) \leqslant \|f\|_\infty$ for every $x \in \mathbb{R}^n$ and $\|Mf\|_\infty \leqslant \|f\|_\infty$. □

Another way to state the previous lemma is that $M : L^\infty(\mathbb{R}^n) \to L^\infty(\mathbb{R}^n)$ is a bounded operator. As we have seen before, $f \in L^1(\mathbb{R})$ does not imply that $Mf \in L^1(\mathbb{R})$ and thus the Hardy-Littlewood maximal operator is not bounded in $L^1(\mathbb{R}^n)$. We give another example of this phenomenon.

Example 2.11. Let $r > 0$. Then there are constants $c_1 = c_1(n)$ and $c_2 = c_2(n)$ such that

$$\frac{c_1 r^n}{(|x|+r)^n} \leqslant M(\chi_{B(0,r)})(x) \leqslant \frac{c_2 r^n}{(|x|+r)^n}$$

for every $x \in \mathbb{R}^n$ (exercise). Since these functions do not belong to $L^1(\mathbb{R}^n)$, we see that the Hardy-Littlewood maximal operator does not map $L^1(\mathbb{R}^n)$ to $L^1(\mathbb{R}^n)$.

Next we show even a stronger result that $Mf \notin L^1(\mathbb{R}^n)$ for every nontrivial $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

Remark 2.12. $Mf \in L^1(\mathbb{R}^n)$ implies $f = 0$.

Reason. Let $r > 0$ and let $x \in \mathbb{R}^n$ such that $|x| \geqslant r$. Then

$$\begin{aligned} Mf(x) &\geqslant \frac{1}{|B(x,2|x|)|}\int_{B(x,2|x|)} |f(y)|\,dy \\ &\geqslant \frac{1}{|B(0,2|x|)|}\int_{B(0,r)} |f(y)|\,dy \\ &\quad (B(0,r) \subset B(x,2|x|),\ |y| < r \Longrightarrow |y-x| \leqslant |y| + |x| < r + |x| \leqslant 2|x|) \\ &= \frac{c}{|x|^n}\int_{B(0,r)} |f(y)|\,dy. \end{aligned}$$

If $f \neq 0$, we choose $r > 0$ large enough that

$$\int_{B(0,r)} |f(y)|\,dy > 0.$$

Then $Mf(x) \geqslant c/|x|^n$ for every $x \in \mathbb{R}^n \setminus B(0,r)$. Since $c/|x|^n \notin L^1(\mathbb{R}^n \setminus B(0,r))$ we conclude that $Mf \notin L^1(\mathbb{R}^n)$. This is a contradiction and thus $f = 0$ almost everywhere. ■

The remark above shows that the maximal function is essentially never in L^1, but the essential issue for this is what happens far away from the origin. The next example shows that the maximal function does not need to be even locally in L^1.

Example 2.13. Let $f : \mathbb{R} \to \mathbb{R}$,

$$f(x) = \frac{\chi_{(0,1/2)}(x)}{x(\log x)^2}.$$

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

$$\int_{\mathbb{R}} |f(x)|\,dx = \int_0^{1/2} \frac{1}{x(\log x)^2}\,dx = \Big|_0^{1/2} -\frac{1}{\log x} < \infty.$$

For $0 < x < 1/2$, we have

$$\begin{aligned} Mf(x) &\geqslant \frac{1}{2x}\int_0^{2x} f(y)\,dy \geqslant \frac{1}{2x}\int_0^{x} f(y)\,dy \\ &= \frac{1}{2x}\int_0^{x} \frac{1}{y(\log y)^2}\,dy = \frac{1}{2x}\Big|_0^{x} -\frac{1}{\log y} \\ &= -\frac{1}{2x\log x} \notin L^1((0,1/2)). \end{aligned}$$

Thus $Mf \notin L^1_{\text{loc}}(\mathbb{R})$.

After these considerations, the situation for L^1 boundedness looks rather hopeless. However, there is a substituting result, which says that if $f \in L^1$, then Mf belongs to a weakened version of L^1.

Definition 2.14. A measurable function $f : \mathbb{R}^n \to [-\infty,\infty]$ belongs to weak $L^1(\mathbb{R}^n)$, if there exists a constant c, $0 \leqslant c < \infty$, such that

$$|\{x \in \mathbb{R}^n : |f(x)| > \lambda\}| \leqslant \frac{c}{\lambda} \quad \text{for every} \quad \lambda > 0.$$

Remarks 2.15:

(1) $L^1(\mathbb{R}^n) \subset$ weak $L^1(\mathbb{R}^n)$.

Reason. By Chebyshev's inequality

$$\begin{aligned} |\{x \in \mathbb{R}^n : |f(x)| > \lambda\}| &\leqslant \frac{1}{\lambda}\int_{\{x\in\mathbb{R}^n : |f(x)|>\lambda\}} |f(y)|\,dy \\ &\leqslant \frac{1}{\lambda}\|f\|_1 \quad \text{for every} \quad \lambda > 0. \end{aligned}$$ ■

(2) Weak $L^1(\mathbb{R}^n) \not\subset L^1(\mathbb{R}^n)$.

Reason. Let $f : \mathbb{R}^n \to [0,\infty]$, $f(x) = |x|^{-n}$. Then $f \notin L^1(\mathbb{R}^n)$, but

$$\begin{aligned} |\{x \in \mathbb{R}^n : |f(x)| > \lambda\}| &= |B(0,\lambda^{-\frac{1}{n}})| \\ &= \Omega_n(\lambda^{-\frac{1}{n}})^n = \Omega_n\lambda^{-1} \quad \text{for every} \quad \lambda > 0. \end{aligned}$$

Here $\Omega_n = |B(0,1)|$. Thus f belongs to weak $L^1(\mathbb{R}^n)$. ■

The next goal is to show that the Hardy-Littlewood maximal operator maps L^1 to weak L^1. The proof is based on the extremely useful covering theorem.

Theorem 2.16 (Vitali covering theorem). Let $\mathscr{F}$ be a collection of open balls B such that

$$\operatorname{diam}\Bigg(\bigcup_{B\in\mathscr{F}} B\Bigg) < \infty.$$

Then there is a countable (or finite) subcollection of pairwise disjoint balls $B(x_i, r_i) \in \mathscr{F}$, $i = 1, 2, \ldots$, such that

$$\bigcup_{B\in\mathscr{F}} B \subset \bigcup_{i=1}^{\infty} B(x_i, 5r_i).$$

THE MORAL: Let A be a bounded subset of $\mathbb{R}^n$ and suppose that for every $x \in A$ there is a ball $B(x, r_x)$ with the radius $r_x > 0$ possibly depending on the point x. We would like to have a countable subcollection of pairwise disjoint balls $B(x_i, r_i)$, $i = 1, 2, \ldots$, which covers the union of the original balls. In general, this is not possible, if we do not expand the balls. Thus

$$\begin{aligned}
|A| &\leqslant \Bigg|\bigcup_{x\in A} B(x, r_x)\Bigg| \leqslant \Bigg|\bigcup_{i=1}^{\infty} B(x_i, 5r_i)\Bigg| \leqslant \sum_{i=1}^{\infty} |B(x_i, 5r_i)| \\
&= 5^n \sum_{i=1}^{\infty} |B(x_i, r_i)| = 5^n \Bigg|\bigcup_{i=1}^{\infty} B(x_i, r_i)\Bigg| \leqslant 5^n \Bigg|\bigcup_{x\in A} B(x, r_x)\Bigg|.
\end{aligned}$$

Note the measure of A can be estimated by the measure of the union of the balls and the measures of $\bigcup_{x\in A} B(x, r_x)$ and $\bigcup_{i=1}^{\infty} B(x_i, r_i)$ are comparable.

THE STRATEGY OF THE PROOF: The greedy principle: The balls are selected inductively by taking the largest ball with the required properties that has not been chosen earlier.

Proof. Assume that $B(x_1, r_1), \ldots, B(x_{i-1}, r_{i-1}) \in \mathscr{F}$ have been selected. Define

$$d_i = \sup\left\{ r : B(x, r) \in \mathscr{F} \quad \text{and} \quad B(x, r) \cap \bigcup_{j=1}^{i-1} B(x_j, r_j) = \emptyset \right\}.$$

Observe that $d_i < \infty$, since $\sup\limits_{B(x,r)\in\mathscr{F}} r < \infty$. If there are no balls $B(x, r) \in \mathscr{F}$ such that

$$B(x, r) \cap \bigcup_{j=1}^{i-1} B(x_j, r_j) = \emptyset,$$

the process terminates and we have selected the balls $B(x_1, r_1), \ldots, B(x_{i-1}, r_{i-1})$. Otherwise, we choose $B(x_i, r_i) \in \mathscr{F}$ such that

$$r_i > \frac{1}{2} d_i \quad \text{and} \quad B(x_i, r_i) \cap \bigcup_{j=1}^{i-1} B(x_j, r_j) = \emptyset.$$

We can also choose the first ball $B(x_1, r_1)$ in this way.

The selected balls are pairwise disjoint. Let $B \in \mathscr{F}$ be an arbitrary ball in the collection $\mathscr{F}$. Then $B = B(x,r)$ intersects at least one of the selected balls $B(x_1,r_1), B(x_2,r_2),\dots$, since otherwise $B(x,r)\cap B(x_i,r_i) = \emptyset$ for every $i = 1,2,\dots$ and, by the definition of d_i, we have $d_i \geqslant r$ for every $i = 1,2,\dots$. This implies

$$r_i > \frac{1}{2}d_i \geqslant \frac{1}{2}r > 0 \quad \text{for every} \quad i = 1,2,\dots,$$

and by the fact that the balls are pairwise disjoint, we have

$$\left|\bigcup_{i=1}^{\infty} B(x_i,r_i)\right| = \sum_{i=1}^{\infty} |B(x_i,r_i)| = \infty.$$

This is impossible, since $\bigcup_{i=1}^{\infty} B(x_i,r_i)$ is bounded and thus $\left|\bigcup_{i=1}^{\infty} B(x_i,r_i)\right| < \infty$.

Since $B(x,r)$ intersects some ball $B(x_i,r_i)$, $i = 1,2,\dots$, there is a smallest index i such that $B(x,r)\cap B(x_i,r_i) \neq \emptyset$. This implies

$$B(x,r)\cap \bigcup_{j=1}^{i-1} B(x_j,r_j) = \emptyset$$

and by the selection process $r \leqslant d_i < 2r_i$. Since $B(x,r)\cap B(x_i,r_i) \neq \emptyset$ and $r \leqslant 2r_i$, we have $B(x,r) \subset B(x_i,5r_i)$.

Reason. Let $z \in B(x,r)\cap B(x_i,r_i)$ and $y \in B(x,r)$. Then

$$|y - x_i| \leqslant |y - z| + |z - x_i| \leqslant 2r + r_i \leqslant 5r_i. \qquad \blacksquare$$

This completes the proof. □

Remarks 2.17:

(1) The factor 5 in the theorem is not optimal. In fact, the same proof shows that this factor can be replaced with 3. A slight modification of the argument shows that $2+\varepsilon$ for any $\varepsilon > 0$ will do. To obtain this, choose $B(x_i,r_i) \in \mathscr{F}$ such that

$$r_i > \frac{1}{1+\varepsilon}d_i$$

in the proof above. This is the optimal result, since 2 does not work in general (exercise).

(2) A similar covering theorem holds true for cubes as well.

(3) Some kind of boundedness assumption is needed in the Vitali covering theorem.

Reason. Let $B(0,i)$, $i = 1,2,\dots$. Since all balls intersect each other, the only subfamily of pairwise disjoint balls consists of one single ball $B(0,i)$ and the enlarged ball $B(0,5i)$ does not cover $\bigcup_{i=1}^{\infty} B(0.i) = \mathbb{R}^n$. $\blacksquare$

Theorem 2.18 (Hardy-Littlewood I). Let $f \in L^1(\mathbb{R}^n)$. Then

$$|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leqslant \frac{5^n}{\lambda}\|f\|_1 \quad \text{for every} \quad \lambda > 0.$$

THE MORAL: The Hardy-Littlewood maximal operator maps L^1 to weak L^1. It is said that the Hardy-Littlewood maximal operator is of weak type $(1,1)$.

Proof. Let $A_\lambda = \{x \in \mathbb{R}^n : Mf(x) > \lambda\}$, $\lambda > 0$. For every $x \in A_\lambda$ there exists $r_x > 0$ such that

$$\frac{1}{|B(x,r_x)|}\int_{B(x,r_x)} |f(y)|\,dy > \lambda \tag{2.1}$$

We would like to apply the Vitali covering theorem, but the set $\bigcup_{x\in A_\lambda} B(x,r_x)$ is not necessarily bounded. To overcome this problem, we consider the sets $A_\lambda \cap B(0,k)$, $k = 1,2,\ldots$. Let $\mathscr{F}$ be the collection of balls for which (2.1) and $x \in A_\lambda \cap B(0,k)$. If $B(x,r_x) \in \mathscr{F}$, then

$$\Omega_n r_x^n = |B(x,r_x)| < \frac{1}{\lambda}\int_{B(x,r_x)} |f(y)|\,dy \leqslant \frac{1}{\lambda}\|f\|_1,$$

so that

$$\operatorname{diam}\left(\bigcup_{x\in A_\lambda\cap B(0,k)} B(x,r_x)\right) < \infty.$$

By the Vitali covering theorem, we obtain pairwise disjoint balls $B(x_i,r_i)$, $i = 1,2,\ldots$, such that

$$A_\lambda \cap B(0,k) \subset \bigcup_{i=1}^{\infty} B(x_i,5r_i).$$

This implies

$$\begin{aligned}
|A_\lambda \cap B(0,k)| &\leqslant \left|\bigcup_{i=1}^{\infty} B(x_i,5r_i)\right| \leqslant \sum_{i=1}^{\infty} |B(x_i,5r_i)| = 5^n \sum_{i=1}^{\infty} |B(x_i,r_i)| \\
&\leqslant \frac{5^n}{\lambda}\sum_{i=1}^{\infty}\int_{B(x_i,r_i)} |f(y)|\,dy = \frac{5^n}{\lambda}\int_{\bigcup_{i=1}^{\infty} B(x_i,r_i)} |f(y)|\,dy \leqslant \frac{5^n}{\lambda}\|f\|_1.
\end{aligned}$$

Finally,

$$|A_\lambda| = \lim_{k\to\infty} |A_\lambda \cap B(0,k)| \leqslant \frac{5^n}{\lambda}\|f\|_1.$$

□

Remark 2.19. $f \in L^1(\mathbb{R}^n)$ implies $Mf < \infty$ almost everywhere in $\mathbb{R}^n$.

Reason.

$$|\{x \in \mathbb{R}^n : Mf(x) = \infty\}| \leqslant |\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| \leqslant \frac{5^n}{\lambda}\|f\|_1 \to 0 \quad \text{as} \quad \lambda \to \infty.$$

■

The next goal is to show that the Hardy-Littlewood maximal operator maps L^p to L^p if $p > 1$. We recall the following Cavalieri's principle.

Lemma 2.20. Assume that μ is an outer measure, $A \subset \mathbb{R}^n$ is μ-measurable set and $f : A \to [-\infty,\infty]$ is a μ-measurable function. Then

$$\int_A |f|^p\,d\mu = p\int_0^{\infty} \lambda^{p-1}\mu(\{x \in A : |f(x)| > \lambda\})\,d\lambda, \quad 0 < p < \infty.$$

Proof.

$$
\begin{aligned}
\int_A |f|^p \, d\mu &= \int_{\mathbb{R}^n} \chi_A(x) p \int_0^{|f(x)|} \lambda^{p-1} \, d\lambda \, d\mu(x) \\
&= p \int_{\mathbb{R}^n} \int_0^\infty \chi_A(x) \chi_{[0,|f(x)|)}(\lambda) \lambda^{p-1} \, d\lambda \, d\mu(x) \\
&= p \int_0^\infty \int_{\mathbb{R}^n} \chi_A(x) \chi_{[0,|f(x)|)}(\lambda) \lambda^{p-1} \, d\mu(x) \, d\lambda \qquad \text{(Fubini's theorem)} \\
&= p \int_0^\infty \lambda^{p-1} \int_{\mathbb{R}^n} \chi_A(x) \chi_{\{x \in \mathbb{R}^n : |f(x)| > \lambda\}}(x) \, d\mu(x) \, d\lambda \\
&= p \int_0^\infty \lambda^{p-1} \mu(\{x \in A : |f(x)| > \lambda\}) \, d\lambda.
\end{aligned}
$$

□

Remark 2.21. More generally, if $\varphi : [0,\infty) \to [0,\infty)$ is a nondecreasing continuously differentiable function with $\varphi(0) = 0$, then

$$
\int_A \varphi \circ |f| \, d\mu = \int_0^\infty \varphi'(\lambda) \mu(\{x \in A : |f(x)| > \lambda\}) \, d\lambda.
$$

(Exercise)

Now we are ready for the Hardy-Littlewood maximal function theorem.

Theorem 2.22 (Hardy-Littlewood II). Let $f \in L^p(\mathbb{R}^n)$, $1 < p \leqslant \infty$. Then $Mf \in L^p(\mathbb{R}^n)$ and there exists $c = c(n,p)$ such that $\|Mf\|_p \leqslant c\|f\|_p$.

THE MORAL: The Hardy-Littlewood maximal operator maps L^p to L^p if $p > 1$. It is said that the Hardy-Littlewood maximal operator is of strong type (p,p).

WARNING:The result is not true $p = 1$. Then we only have the weak type estimate.

Proof. Let $f = f_1 + f_2$, where $f_1 = f\chi_{\{|f| > \lambda/2\}}$, that is,

$$
f_1(x) = \begin{cases} f(x), & |f(x)| > \lambda/2, \\ 0, & |f(x)| \leqslant \lambda/2. \end{cases}
$$

Then $|f_1(x)| > \lambda/2$ if $|f(x)| > \lambda/2$ and thus

$$
\begin{aligned}
\int_{\mathbb{R}^n} |f_1(x)| \, dx &= \int_{\{x \in \mathbb{R}^n : |f(x)| > \lambda/2\}} |f_1(x)|^p |f_1(x)|^{1-p} \, dx \\
&\leqslant \left(\frac{\lambda}{2}\right)^{1-p} \|f\|_p^p < \infty.
\end{aligned}
$$

This shows that $f_1 \in L^1(\mathbb{R}^n)$. On the other hand, $|f_2(x)| \leqslant \lambda/2$ for every $x \in \mathbb{R}^n$, which implies $\|f_2\|_\infty \leqslant \lambda/2$ and $f_2 \in L^\infty(\mathbb{R}^n)$. Thus every L^p function can be represented as a sum of an L^1 function and an L^∞ function. By Lemma 2.10, we have

$$
\|Mf_2\|_\infty \leqslant \|f_2\|_\infty \leqslant \frac{\lambda}{2}
$$

From this we conclude using Lemma 2.9 that

$$Mf(x) = M(f_1 + f_2)(x) \leqslant Mf_1(x) + Mf_2(x) \leqslant Mf_1(x) + \frac{\lambda}{2}$$

for almost every $x \in \mathbb{R}^n$ and thus $Mf(x) > \lambda$ implies $Mf_1(x) > \lambda/2$. It follows that

$$\begin{aligned}
|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| &\leqslant \left|\left\{x \in \mathbb{R}^n : Mf_1(x) > \frac{\lambda}{2}\right\}\right| \\
&\leqslant \frac{2\cdot 5^n}{\lambda}\|f_1\|_1 \qquad \text{(Theorem 2.18)} \\
&= \frac{2\cdot 5^n}{\lambda}\int_{\{x\in\mathbb{R}^n : |f(x)|>\lambda/2\}} |f(x)|\,dx \quad \text{for every} \quad \lambda > 0
\end{aligned}$$

By Cavalieri's principle

$$\begin{aligned}
\int_{\mathbb{R}^n} |Mf|^p\,dx &= p\int_0^\infty \lambda^{p-1}|\{x \in \mathbb{R}^n : Mf(x) > \lambda\}|\,d\lambda \\
&\leqslant p\cdot 2\cdot 5^n \int_0^\infty \lambda^{p-2}\int_{\{x\in\mathbb{R}^n : |f(x)|>\lambda/2\}} |f(x)|\,dx\,d\lambda \\
&= p\cdot 2\cdot 5^n \int_{\mathbb{R}^n} |f(x)|\int_0^{2|f(x)|} \lambda^{p-2}\,d\lambda\,dx \\
&\qquad \text{(as in the proof of Lemma 2.20)} \\
&= \frac{p\cdot 2\cdot 5^n}{p-1}\int_{\mathbb{R}^n} |f(x)||2f(x)|^{p-1}\,dx \\
&= \frac{p\cdot 2^p\cdot 5^n}{p-1}\int_{\mathbb{R}^n} |f(x)|^p\,dx.
\end{aligned}$$

This completes the proof. □

Remarks 2.23:

(1) The proof above gives

$$\|Mf\|_p \leqslant 2\left(\frac{p5^n}{p-1}\right)^{1/p}\|f\|_p, \quad 1 < p < \infty$$

for the operator norm of $M : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$. Note that it blows up as $p \to 1$ and converges to 2 as $p \to \infty$.

It can be shown that that the constant in the strong type (p,p) inequality for $p > 1$ can be chosen to be independent of the dimension n, see E. M. Stein and J.-O. Strömberg, *Behavior of maximal functions in* $\mathbb{R}^n$ *for large* n, Ark. Mat. **21** (1983), 259–269. However, it is an open question whether the constant in the weak type (1,1) estimate is independent of the dimension.

(2) As a byproduct of the proof we get the following useful result. Let $1 \leqslant p < r < q \leqslant \infty$. Then for every $f \in L^r(\mathbb{R}^n)$ there exist $g \in L^p(\mathbb{R}^n)$ and $h \in L^q(\mathbb{R}^n)$ such that $f = g + h$. Hint: $g = f\chi_{\{|f|>1\}}$.

(3) The proof above is a special case of the Marcinkiewicz interpolation theorem, which applies to more general operators as well. In this case, we interpolate between the weak type (1,1) estimate and the strong type (∞,∞) estimate.

2.4 Lebesgue's differentiation theorem

Lebesgue differentiation theorem is a remarkable result, which shows that a quantitative weak type estimate for the maximal function implies almost everywhere convergence of integral averages using the fact that the convergence is clear for a dense class of continuous functions.

Theorem 2.24. Assume $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then

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

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

Remark 2.25. In particular, it follows that

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

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

Reason.

$$\begin{aligned}\left|\frac{1}{|B(x,r)|}\int_{B(x,r)} f(y)\,dy - f(x)\right| &= \left|\frac{1}{|B(x,r)|}\int_{B(x,r)} (f(y)-f(x))\,dy\right| \\ &\leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)-f(x)|\,dy \;\to 0\end{aligned}$$

for almost every $x \in \mathbb{R}^n$. ■

Note that this implies

$$\begin{aligned}|f(x)| &= \lim_{r\to 0} \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy \\ &\leqslant \sup_{r>0} \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy = Mf(x)\end{aligned}$$

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

THE MORAL: A locally integrable function is a limit of the integral averages at almost every point. Observe, that Lebesgue's differentiation tells that the limit of the integral averages exists and that it coincides with the function almost everywhere. This gives a passage from average information to pointwise information.

Proof. We may assume that $f \in L^1(\mathbb{R}^n)$, since the theorem is local. Indeed, we may consider the functions $f_i = f\chi_{B(0,i)}$, $i = 1,2,\ldots$. Define a infinitesimal version of the Hardy-Littlewood maximal function as

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

We shall show that $f^*(x) = 0$ for almost every $x \in \mathbb{R}^n$. The proof is divided into six steps.

(1) Clearly $f^* \geqslant 0$.

(2) $(f+g)^* \leqslant f^* + g^*$.

Reason.

$$\begin{aligned}
&\frac{1}{|B(x,r)|}\int_{B(x,r)} |(f+g)(y)-(f+g)(x)|\,dy \\
&\qquad \leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)-f(x)|\,dy + \frac{1}{|B(x,r)|}\int_{B(x,r)} |g(y)-g(x)|\,dy. \qquad \blacksquare
\end{aligned}$$

(3) If g is continuous at x, then $g^*(x) = 0$.

Reason. For every $\varepsilon > 0$ there exists $\delta > 0$ such that $|g(y)-g(x)| < \varepsilon$ whenever $|x-y| < \delta$. This implies

$$\frac{1}{|B(x,r)|}\int_{B(x,r)} |g(y)-g(x)|\,dy < \varepsilon, \quad \text{if} \quad 0 < r \leqslant \delta.$$

■

(4) If $g \in C(\mathbb{R}^n)$, then $(f-g)^* = f^*$.

Reason. By (3), we have

$$(f-g)^* \leqslant f^* + (-g)^* = f^* \quad \text{and} \quad f^* \leqslant (f-g)^* + g^* = (f-g)^*,$$

so that the equality holds. ■

(5) $f^* \leqslant Mf + |f|$.

Reason.

$$\begin{aligned}
\frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)-f(x)|\,dy &\leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)} (|f(y)|+|f(x)|)\,dy \\
&\leqslant \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(y)|\,dy + |f(x)| \\
&\leqslant Mf(x) + |f(x)|. \qquad \blacksquare
\end{aligned}$$

(6) If $f^*(x) > \lambda$, then by (5) we have $Mf(x) + |f(x)| > \lambda$, from which we conclude that $Mf(x) > \lambda/2$ or $|f(x)| > \lambda/2$. Thus

$$\begin{aligned}
|\{x \in \mathbb{R}^n : f^*(x) > \lambda\}| &\leqslant \left|\left\{x \in \mathbb{R}^n : Mf(x) > \frac{\lambda}{2}\right\}\right| + \left|\left\{x \in \mathbb{R}^n : |f(x)| > \frac{\lambda}{2}\right\}\right| \\
&\leqslant \frac{2\cdot 5^n}{\lambda}\|f\|_1 + \frac{2}{\lambda}\|f\|_1 \qquad \text{(Theorem 2.18 and Chebyshev)} \\
&= \frac{2(5^n+1)}{\lambda}\|f\|_1
\end{aligned}$$

Finally, we are ready to prove the theorem. Recall from the measure and integration theory that compactly supported continuous functions are dense in $L^1(\mathbb{R}^n)$. Thus for every $\varepsilon > 0$ there exists $g \in C_0(\mathbb{R}^n)$ such that $\|f-g\|_1 < \varepsilon$. Then

$$\begin{aligned}
|\{x \in \mathbb{R}^n : f^*(x) > \lambda\}| &= |\{x \in \mathbb{R}^n : (f-g)^*(x) > \lambda\}| \qquad \text{(Property (4))} \\
&\leqslant \frac{2(5^n+1)}{\lambda}\|f-g\|_1 \qquad \text{(Property (6))} \\
&< \frac{2(5^n+1)}{\lambda}\varepsilon.
\end{aligned}$$

Letting $\varepsilon \to 0$, we conclude that $|\{x \in \mathbb{R}^n : f^*(x) > \lambda\}| = 0$ for every $\lambda > 0$. It follows that

$$\begin{aligned}
|\{x \in \mathbb{R}^n : f^*(x) > 0\}| &= \left|\bigcup_{i=1}^{\infty}\left\{x \in \mathbb{R}^n : f^*(x) > \frac{1}{i}\right\}\right| \\
&\leqslant \sum_{i=1}^{\infty} \underbrace{\left|\left\{x \in \mathbb{R}^n : f^*(x) > \frac{1}{i}\right\}\right|}_{=0} = 0.
\end{aligned}$$

This shows that $f^*(x) \leqslant 0$ for almost every $x \in \mathbb{R}^n$ and (1) implies $f^*(x) = 0$ for almost every $x \in \mathbb{R}^n$. □

Definition 2.26. A point $x \in \mathbb{R}^n$ is a Lebesgue point of $f \in L^1_{\text{loc}}(\mathbb{R}^n)$, if there exists $a \in \mathbb{R}$ such that

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

T H E M O R A L : The Lebesgue differentiation theorem asserts that almost every point is a Lebesgue point for a locally integrable function. Thus locally integrable function can be defined pointwise almost everywhere.

Remarks 2.27:

(1) We would like to define the Lebesgue point so that a is replaced with $f(x)$, but there is a problem with this definition since the equivalence class of f is defined only up to a set of measure zero. If $f = g$ almost everywhere, the functions have the same Lebesgue points. Thus the notion of a Lebesgue point is independent of the representative in the equivalence class in $L^1_{\text{loc}}(\mathbb{R}^n)$.

(2) If x is a Lebesgue point of f, then

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

Thus we may uniquely define the pointwise value of f by the above limit at a Lebesgue point.

(3) Whether x is a Lebesgue point of f is completely independent of the value $f(x)$. In fact, the function f does not even need to be defined at x. By the Lebesgue differentiation theorem, almost every point $x \in \mathbb{R}^n$ is a Lebesgue point of $f \in L^1_{\text{loc}}(\mathbb{R}^n)$. Moreover, if f is a specific function in the equivalence class in $L^1_{\text{loc}}(\mathbb{R}^n)$, then for almost every x the number a is $f(x)$.

Example 2.28. Let $f:\mathbb{R}\to\mathbb{R}$ be the Heaviside function

$$f(x)=\begin{cases}1, & x>0,\\ \frac{1}{2}, & x=0,\\ 0, & x<0.\end{cases}$$

Then

$$\lim_{r\to 0}\frac{1}{2r}\int_{x-r}^{x+r}f(y)\,dy=f(x)\quad\text{for every}\quad x\in\mathbb{R},$$

but 0 is not a Lebesgue point of f.

Reason.

$$\begin{aligned}\frac{1}{2r}\int_{-r}^{r}|f(y)-a|\,dy&=\frac{1}{2r}\int_{-r}^{0}|a|\,dy+\frac{1}{2r}\int_{0}^{r}|1-a|\,dy\\&=\frac{1}{2}|a|+\frac{1}{2}|1-a|\neq 0\quad\text{for every}\quad a\in\mathbb{R},r>0.\end{aligned}$$

■

Next we remark that the use of balls is not crucial in the Lebesgue differentiation theorem. The theory of maximal functions can be done with cubes instead of balls, for example. As we shall see, the geometry of the sets does not play role here.

Definition 2.29. A sequence of measurable sets A_i, $i=1,2,\dots$, converges regularly to a point $x\in\mathbb{R}^n$, if there exist a constant $c>0$ and a sequence of positive numbers r_i, $i=1,2,\dots$, such that

(1) $A_i\subset B(x,r_i)$, $i=1,2,\dots$,

(2) $\lim_{i\to\infty}r_i=0$ and

(3) $|A_i|\leqslant|B(x,r_i)|\leqslant c|A_i|$, $i=1,2,\dots$.

T H E M O R A L : The conditions (1) and (2) ensure that the sets A_i converge to x. The condition (3) ensures that the convergence is not too fast with respect to the Lebesgue measure: the volume of each A_i is at least certain percentage of the volume of $B(x,r_i)$. Note that x does not have to belong to the sets A_i.

Examples 2.30:

(1) Let

$$Q(x,l)=\left\{y\in\mathbb{R}^n:|y_i-x_i|<\frac{l}{2},\ i=1,\dots,n\right\}$$

be an open cube with the center $x\in\mathbb{R}^n$ and the side length $l>0$.

C L A I M : $Q\left(x,\frac{2}{\sqrt{n}}r\right)\subset B(x,r)$.

Reason. Let $y\in Q(x,\frac{2}{\sqrt{n}}r)$. Then $|y_i-x_i|<\frac{r}{\sqrt{n}}$, $i=1,\dots,n$, which implies

$$|y-x|=\left(\sum_{i=1}^{n}|y_i-x_i|^2\right)^{1/2}<\left(n\cdot\left(\frac{r}{\sqrt{n}}\right)^2\right)^{1/2}=r.$$

Thus $y\in B(x,r)$. ■

CLAIM: $|B(x,r)| = c\left|Q\left(x,\frac{2}{\sqrt{n}}r\right)\right|$.

Reason.

$$
\begin{aligned}
|B(x,r)| &= \frac{|B(x,r)|}{|Q(x,\frac{2}{\sqrt{n}}r)|}\left|Q\left(x,\frac{2}{\sqrt{n}}r\right)\right| \\
&= \frac{\Omega_n r^n}{(\frac{2}{\sqrt{n}})^n r^n}\left|Q\left(x,\frac{2}{\sqrt{n}}r\right)\right| \qquad (\Omega_n = |B(0,1)|) \\
&= c\left|Q\left(x,\frac{2}{\sqrt{n}}r\right)\right|, \quad c = \frac{\Omega_n n^{n/2}}{2^n}.
\end{aligned}
$$

■

Thus the the cubes $Q(x,\frac{2}{\sqrt{n}}r_i)$ converge regularly to x if $r_i \to 0$ as $i \to \infty$.

(2) Let $A \subset B(0,1)$ be arbitrary measurable set with $|A| > 0$ and denote

$$A_r(x) = x + rA = \{y \in \mathbb{R}^n : y = x + rz,\ z \in A\}.$$

Then $A_r(x) \subset x + rB(0,1) = B(x,r)$ and

$$
\begin{aligned}
|B(x,r)| &= \frac{|B(x,r)|}{|A_r(x)|}|A_r(x)| \\
&= \frac{r^n|B(0,1)|}{r^n|A|}|A_r(x)| \\
&= c|A_r(x)|, \quad c = \frac{|B(0,1)|}{|A|}
\end{aligned}
$$

Thus the the sets A_{r_i} converge regularly to x if $r_i \to 0$ as $i \to \infty$. This means that we can construct a sequence that converges regularly from an arbitrary set $A \subset B(0,1)$ with $|A| > 0$.

For example, if $A = B(0,1)\setminus B(0,1/2)$, then $A_r(x) = B(x,r)\setminus B(x,r/2)$ and $x \notin A_r(x)$ for any $r > 0$.

Theorem 2.31. Assume that $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and let x be a Lebesgue point of f. If the sequence A_i, $i = 1,2,\dots$, converges regularly to x, then

$$\lim_{i\to\infty}\frac{1}{|A_i|}\int_{A_i}|f(y)-f(x)|\,dy = 0.$$

THE MORAL: The Lebesgue differentiation theorem holds for any regularly converging sets.

Proof.

$$\frac{1}{|A_i|}\int_{A_i}|f(y)-f(x)|\,dy \leqslant \frac{c}{|B(x,r_i)|}\int_{B(x,r_i)}|f(y)-f(x)|\,dy \to 0 \quad \text{as} \quad i\to\infty.$$

□

Remark 2.32. The convergence of the previous theorem is valid. Assume that $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ and let $x \in \mathbb{R}^n$. If for every sequence A_i, $i = 1,2,\dots$, that converges regularly to x, there exists

$$\lim_{i\to\infty}\frac{1}{|A_i|}\int_{A_i}f(y)\,dy,$$

then x is a Lebesgue point of f. (Exercise)

Hint: By interlacing two sequences, show that the limit is independent of the sequence. Then show that we may assume that the limit is zero. Then assume that $r_i \to 0$ and take

$$A_i = B(x,r_i) \cap \{y \in \mathbb{R}^n : f(y) \geqslant 0\} \quad \text{or} \quad A_i = B(x,r_i) \cap \{y \in \mathbb{R}^n : f(y) < 0\}$$

depending on which choice satisfies $|A_i| \geqslant |B(x,r_i)|/2$. Show that

$$\lim_{i\to\infty} \frac{1}{|B(x,r_i)|} \int_{B(x,r_i)} f(y)\,dy = 0.$$

2.5 The fundamental theorem of calculus

As an application of the Lebesgue differentiation theorem, we prove the following theorem of Lebesgue in the one-dimensional case.

Theorem 2.33. Assume that $f \in L^1([a,b])$ and define $F : [a,b] \to \mathbb{R}$,

$$F(x) = \int_{[a,x]} f(y)\,dy.$$

Then $F'(x)$ exists and $F'(x) = f(x)$ for almost every $x \in [a,b]$.

THE MORAL: This is a general version of the fundamental theorem of calculus, which is elementary in the case $f \in C([a,b])$.

Proof. Define $f(x) = 0$ for every $x \in \mathbb{R} \setminus [a,b]$. Let $r_i > 0$ with $\lim_{i\to\infty} r_i = 0$ and denote $A_i = (x, x+r_i)$, $i = 1,2,\dots$. Then the sets A_i converge regularly to x. By Theorem 2.31

$$\lim_{i\to\infty} \frac{F(x+r_i)-F(x)}{r_i} = \lim_{i\to\infty} \frac{1}{r_i} \int_{(x,x+r_i)} f(y)\,dy = f(x)$$

for almost every $x \in \mathbb{R}$. Since the sequence is arbitrary, we conclude that $F'_+(x)$ exists and $F'_+(x) = f(x)$ for almost every $x \in \mathbb{R}$.

Similarly, by choosing $A_i = (x - r_i, x)$, $i = 1,2,\dots$, we obtain

$$\lim_{i\to\infty} \frac{F(x-r_i)-F(x)}{r_i} = f(x)$$

and $F'_-(x) = f(x)$ for almost every $x \in \mathbb{R}$. Therefore $F'(x)$ exists and $F'(x) = f(x)$ for almost every $x \in [a,b]$. □

Remark 2.34. Assume that $f \in L^1([a,b])$ and define $F : [a,b] \to \mathbb{R}$,

$$F(x) = F(a) + \int_{[a,x]} f(y)\,dy.$$

Then $F'(x) = f(x)$ for almost every $x \in [a,b]$ and thus

$$F(x) = F(a) + \int_{[a,x]} F'(y)\,dy. \tag{2.2}$$

PROBLEM: What do we have to assume about the function F to guarantee that (2.2) holds?

(1) If $F \in C^1([a,b])$, then (2.2) holds.

(2) If $F = \chi_{[-1,1]}$ then $F' = 0$ almost everywhere in $\mathbb{R}$, but (2.2) does not hold.

(3) It is not enough that F is differentiable everywhere.

Reason. Let $F : \mathbb{R} \to \mathbb{R}$,

$$F(x) = \begin{cases} x^2 \sin \frac{1}{x}, & x \neq 0 \\ 0, & x = 0. \end{cases}$$

Then $F'(x)$ exists for every $x \in \mathbb{R}$, but $F' \notin L^1(\mathbb{R})$ (exercise). Thus (2.2) does not make sense. ■

(4) It is not enough that $F \in C([a,b])$, $F'(x)$ exists for almost every $x \in [a,b]$ and $F' \in L^1([a,b])$.

Reason. For the Cantor-Lebesgue function (see Measure and Integral)

$$F(1) = 1 \neq 0 = F(0) + \int_{[0,1]} F'(y)\,dy.$$ ■

THE FINAL ANSWER: The formula (2.2) defines an important class of functions: A function $F : [a,b] \to \mathbb{R}$ is absolutely continuous if there exists $f \in L^1([a,b])$ such that

$$F(x) = F(a) + \int_{[a,x]} f(y)\,dy$$

for every $x \in [a,b]$. It follows that $f(x) = F'(x)$ for almost every $x \in [a,b]$.

2.6 Points of density

We discuss a special case of the Lebesgue differentiability theorem. Let $A \subset \mathbb{R}^n$ a measurable set and consider $f = \chi_A$. By the Lebesgue differentation theorem

$$\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} \chi_A(y)\,dy = \lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = \chi_A(x)$$

for almost every $x \in \mathbb{R}^n$. In particular,

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1 \quad \text{for almost every} \quad x \in A$$

and

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 0 \quad \text{for almost every} \quad x \in \mathbb{R}^n \setminus A.$$

Definition 2.35. Let A be an arbitrary subset of $\mathbb{R}^n$. A point $x \in \mathbb{R}^n$ is a point of density of A, if

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1.$$

THE MORAL: Density points are measure theoretic interior points of the set. Loosely speaking, the small balls around x are almost entirely covered by A. The points with zero density belong to the measure theoretic complement of the set. In this case, the small balls around x are almost entirely covered by the complement of A. The Lebesgue differentiation theorem asserts that almost every point of a measurable set is a density point and almost every point of the complement of measurable set is a point of zero density.

Examples 2.36:

(1) Let $I_i = [2^{-(2i+1)}, 2^{-2i}]$, $i = 1,2,\dots$. Then $|I_i| = 2^{-2i} - 2^{-(2i+1)} = 2^{-(2i+1)}$, $i = 1,2,\dots$. Let $A = \bigcup_{i=1}^{\infty} I_i$. Then

$$A \cap B(0, 2^{-2k}) = \bigcup_{i=k}^{\infty} I_i$$

and thus

$$|A \cap B(0,2^{-2k})| = \sum_{i=k}^{\infty} \frac{1}{2^{2i+1}} = \frac{4}{3}\frac{1}{2^{2k+1}}, \quad k = 1,2,\dots.$$

This implies

$$\frac{|A \cap B(0,2^{-2k})|}{|B(0,2^{-2k})|} = \frac{4}{3}\frac{1}{2^{2k+1}} \cdot \frac{2^{2k}}{2} = \frac{1}{3}$$

and

$$\frac{|A \cap B(0,2^{-(2k+1)})|}{|B(0,2^{-(2k+1)})|} = \frac{4}{3}\frac{1}{2^{2k+3}} \cdot \frac{2^{2k+1}}{2} = \frac{1}{6}.$$

Thus the limit

$$\lim_{r\to 0} \frac{|A \cap B(0,r)|}{|B(0,r)|}$$

does not exist and $x = 0$ is not a density point of A.

(2) Let $A = \{x \in \mathbb{R}^2 : |x_i| < 1, \ i = 1,2\}$. Then

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = \begin{cases} 1, & x \in A, \\ \frac{1}{2}, & x \in \partial A \setminus \{(1,1),(-1,1),(-1,-1),(1,-1)\}, \\ \frac{1}{4}, & x \in \{(1,1),(-1,1),(-1,-1),(1,-1)\}, \\ 0, & x \in \mathbb{R}^2 \setminus \overline{A}. \end{cases}$$

(3) Let $A = \{x = re^{i\theta} : r > 0, \ 0 \leqslant \theta \leqslant 2\pi\alpha\}$, $0 < \alpha < 1$. Then

$$\lim_{r\to 0} \frac{|A \cap B(0,r)|}{|B(0,r)|} = \lim_{r\to 0} \frac{2\pi\alpha}{2\pi} = \alpha.$$

Remarks 2.37:

(1) There does not exist a Lebesgue measurable set $A \subset \mathbb{R}^n$ such that

$$|A \cap B(x,r)| = \frac{1}{2}|B(x,r)| \quad \text{for every} \quad x \in A, r > 0.$$

Reason. Assume that there exists such a set A. Note that

$$|A| \geqslant |A \cap B(x,r)| = \frac{1}{2}|B(x,r)| > 0, \quad \text{if} \quad r > 0.$$

By the Lebesgue differentiation theorem

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1$$

for almost every $x \in A$ and thus on a set of positive measure in A. This contradicts with the fact that

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = \frac{1}{2}$$

for every $x \in A$. ■

(2) Let $A \subset \mathbb{R}^n$ be a measurable set. Then

$$|A| > 0 \Longleftrightarrow A \text{ has a Lebesgue point.}$$

Reason. $\boxed{\Longrightarrow}$ By the Lebesgue differentiation theorem the a set of Lebesgue points of $f = \chi_A$ in the set A has positive measure. Thus there exists at least one point with the required property.

$\boxed{\Longleftarrow}$ Assume that there exists $x \in A$ such that

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1.$$

Then for every $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\frac{|A \cap B(x,r)|}{|B(x,r)|} > 1 - \varepsilon \quad \text{when} \quad 0 < r < \delta.$$

This implies

$$|A| \geqslant |A \cap B(x,r)| > (1-\varepsilon)|B(x,r)| > 0, \quad 0 < \varepsilon < 1.$$ ■

(3) If $A \subset \mathbb{R}^n$ is a measurable set such that

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} < 1$$

for every $x \in A$, then $|A| = 0$.

(4) Assume that $\Omega \subset \mathbb{R}^n$ is an open set. If there exists γ, $0 < \gamma \leqslant 1$, such that

$$|\Omega \cap B(x,r)| \geqslant \gamma |B(x,r)| \quad \text{for every} \quad x \in \partial\Omega, r > 0,$$

then $|\partial\Omega| = 0$.

Recall that the complement of a fat Cantor set is an open set whose boundary has positive measure. This shows that the claim above is nontrivial.

Reason. Since $\Omega \subset \mathbb{R}^n$ is open, we have $\partial\Omega \subset \mathbb{R}^n \setminus \Omega$. By the Lebesgue differentiation theorem

$$\lim_{r\to 0} \frac{|\Omega \cap B(x,r)|}{|B(x,r)|} = 0 \quad \text{for almost every} \quad x \in \partial\Omega.$$

On the other hand,

$$\lim_{r\to 0} \frac{|\Omega \cap B(x,r)|}{|B(x,r)|} \geqslant \gamma > 0 \quad \text{for every} \quad x \in \partial\Omega.$$

Thus $|\partial\Omega| = 0$. ■

(5) Let A be an arbitrary subset of $\mathbb{R}^n$. Then

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1 \quad \text{for almost every} \quad x \in A.$$

Note that this holds without the assumption that A is measurable.

Reason. Since the Lebesgue outer measure is Borel regular, there exists a Borel set B such that $A \subset B$ and $|B \setminus A| = 0$. Then for every measurable set E, we have

$$|A \cap E| = |A \cap E| + |(B \setminus A) \cap E| = |B \cap E|,$$

where in the last equality we used the definition of a measurable set. Thus

$$\frac{|A \cap B(x,r)|}{|B(x,r)|} = \frac{|B \cap B(x,r)|}{|B(x,r)|}.$$

Since B is measurable, we have

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = \lim_{r\to 0} \frac{|B \cap B(x,r)|}{|B(x,r)|} = 1 \quad \text{for almost every} \quad x \in B.$$

The claim follows, since $A \subset B$. ■

(6) A set $A \subset \mathbb{R}^n$ is measurable if and only if

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 0 \quad \text{for almost every} \quad x \in \mathbb{R}^n \setminus A.$$

Reason. $\boxed{\Longrightarrow}$ This follows from the Lebesgue differentiation theorem.
$\boxed{\Longleftarrow}$ Assume that

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 0 \quad \text{for almost every} \quad x \in \mathbb{R}^n \setminus A.$$

By (5),

$$\lim_{r\to 0} \frac{|A \cap B(x,r)|}{|B(x,r)|} = 1 \quad \text{for almost every} \quad x \in B.$$

we conclude that $B \setminus A$ is a set of measure zero and thus measurable. This implies that $A = B \setminus (B \setminus A)$ is measurable. ■

2.7 The Sobolev embedding

We begin with considering the one-dimensional case. If $u \in C_0^1(\mathbb{R})$, there exists an interval $[a,b] \subset \mathbb{R}$ such that $u(x) = 0$ for every $x \in \mathbb{R} \setminus [a,b]$. By the fundamental theorem of calculus,

$$u(x) = u(a) + \int_a^x u'(y)\,dy = \int_{-\infty}^x u'(y)\,dy, \tag{2.3}$$

since $u(a) = 0$. On the other hand,

$$0 = u(b) = u(x) + \int_x^b u'(y)\,dy = u(x) + \int_x^\infty u'(y)\,dy,$$

so that

$$u(x) = -\int_x^\infty u'(y)\,dy. \tag{2.4}$$

Equalities (2.3) and (2.4) imply

$$\begin{aligned} 2u(x) &= \int_{-\infty}^x u'(y)\,dy - \int_x^\infty u'(y)\,dy \\ &= \int_{-\infty}^x \frac{u'(y)(x-y)}{|x-y|}\,dy + \int_x^\infty \frac{u'(y)(x-y)}{|x-y|}\,dy \\ &= \int_{\mathbb{R}} \frac{u'(y)(x-y)}{|x-y|}\,dy, \end{aligned}$$

from which it follows that

$$u(x) = \frac{1}{2}\int_{\mathbb{R}} \frac{u'(y)(x-y)}{|x-y|}\,dy \quad \text{for every} \quad x \in \mathbb{R}.$$

Next we extend this to $\mathbb{R}^n$.

Lemma 2.38. If $u \in C_0^1(\mathbb{R}^n)$, then

$$u(x) = \frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{\nabla u(y)\cdot(x-y)}{|x-y|^n}\,dy \quad \text{for every} \quad x \in \mathbb{R}^n,$$

where $\omega_{n-1} = n\Omega_n$ on $(n-1)$-dimensional measure of $\partial B(0,1)$ and

$$\nabla u = \left(\frac{\partial u}{\partial x_1}, \dots, \frac{\partial u}{\partial x_n}\right)$$

is the gradient of u.

Proof. If $x \in \mathbb{R}^n$ and $e \in \partial B(0,1)$, by the fundamental theorem of calculus

$$u(x) = -\int_0^\infty \frac{\partial}{\partial t} u(x+te)\,dt = -\int_0^\infty \nabla u(x+tv)\cdot e\,dt.$$

By the Fubini theorem

$$
\begin{aligned}
\omega_{n-1}u(x) &= u(x)\int_{\partial B(0,1)} 1\,dS(e) \\
&= -\int_{\partial B(0,1)}\int_0^\infty \nabla u(x+te)\cdot e\,dt\,dS(e) \\
&= -\int_0^\infty \int_{\partial B(0,1)} \nabla u(x+te)\cdot e\,dS(e)\,dt \qquad \text{(Fubini)} \\
&= -\int_0^\infty \int_{\partial B(0,t)} \nabla u(x+y)\cdot \frac{y}{t}\,\frac{1}{t^{n-1}}\,dS(y)\,dt \\
&\quad (y=te,\ dS(e)=t^{1-n}\,dS(y)) \\
&= -\int_0^\infty \int_{\partial B(0,t)} \nabla u(x+y)\cdot \frac{y}{|y|^n}\,dS(y)\,dt \\
&= -\int_{\mathbb{R}^n} \frac{\nabla u(x+y)\cdot y}{|y|^n}\,dy \qquad \text{(polar coordinates)} \\
&= -\int_{\mathbb{R}^n} \frac{\nabla u(z)\cdot (z-x)}{|z-x|^n}\,dz \qquad (z=x+y,\ dy=dz) \\
&= \int_{\mathbb{R}^n} \frac{\nabla u(y)\cdot (x-y)}{|x-y|^n}\,dy. \qquad \square
\end{aligned}
$$

By the Cauchy-Schwarz inequality and Lemma 2.38, we have

$$
\begin{aligned}
|u(x)| &\leqslant \frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{|\nabla u(y)||x-y|}{|x-y|^n}\,dy \\
&= \frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{|\nabla u(y)|}{|x-y|^{n-1}}\,dy \\
&= \frac{1}{\omega_{n-1}} I_1(|\nabla u|)(x),
\end{aligned}
$$

where $I_\alpha f$, $0<\alpha<n$, is the Riesz potential

$$
I_\alpha f(x) = \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-\alpha}}\,dy. \tag{2.5}
$$

Lemma 2.39. If $0<\alpha<n$, there exists a constant $c=c(n,\alpha)>0$, such that

$$
\int_{B(x,r)} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy \leqslant cr^\alpha Mf(x)
$$

for every $x\in\mathbb{R}^n$ and $r>0$.

Proof. Let $x \in \mathbb{R}^n$ and denote $A_i = B(x, r2^{-i})$. Then

$$\begin{aligned}
\int_{B(x,r)} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy &= \sum_{i=0}^{\infty} \int_{A_i \setminus A_{i+1}} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy \\
&\leqslant \sum_{i=0}^{\infty} \left(\frac{r}{2^{i+1}}\right)^{\alpha-n} \int_{A_i} |f(y)|\,dy \\
&= \Omega_n \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{\alpha-n} \left(\frac{r}{2^i}\right)^{\alpha} \frac{1}{\Omega_n} \left(\frac{r}{2^i}\right)^{-n} \int_{A_i} |f(y)|\,dy \\
&= \Omega_n \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{\alpha-n} \left(\frac{r}{2^i}\right)^{\alpha} \frac{1}{|A_i|} \int_{A_i} |f(y)|\,dy \\
&= c\,Mf(x) r^{\alpha} \sum_{i=0}^{\infty} \left(\frac{1}{2^{\alpha}}\right)^{i} \\
&= c\,r^{\alpha} Mf(x).
\end{aligned}$$

□

Theorem 2.40 (The Sobolev inequality for the Riesz potentials). Assume that $\alpha > 0$, $1 < p < n$ and $\alpha p < n$. Then there exists a constant $c = c(n, p, \alpha) > 0$, such that for every $f \in Ł(\mathbb{R}^n)$ we have

$$\|I_\alpha f\|_{p^*} \leqslant c\|f\|_p, \quad \text{where} \quad p^* = \frac{pn}{n-\alpha p}.$$

Proof. By Hölder's inequality

$$\int_{\mathbb{R}^n \setminus B(x,r)} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy \leqslant \left(\int_{\mathbb{R}^n \setminus B(x,r)} |f(y)|^p\,dy\right)^{1/p} \left(\int_{\mathbb{R}^n \setminus B(x,r)} |x-y|^{(\alpha-n)p'}\,dy\right)^{1/p'},$$

where

$$\begin{aligned}
\int_{\mathbb{R}^n \setminus B(x,r)} |x-y|^{(\alpha-n)p'}\,dy &= \int_r^{\infty} \int_{\partial B(x,\rho)} |x-y|^{(\alpha-n)p'}\,dS(y)\,d\rho \\
&= \int_r^{\infty} \rho^{(\alpha-n)p'} \underbrace{\int_{\partial B(x,\rho)} 1\,dS(y)}_{=\omega_{n-1}\rho^{n-1}}\,d\rho \\
&= \omega_{n-1} \int_r^{\infty} \rho^{(\alpha-n)p'+n-1}\,d\rho \\
&= \frac{\omega_{n-1}}{(n-\alpha)p'-n} r^{n-(n-\alpha)p'}.
\end{aligned}$$

The exponent can be written in the form

$$n-(n-\alpha)p' = n-(n-\alpha)\frac{p}{p-1} = \frac{\alpha p - n}{p-1},$$

and thus

$$\int_{\mathbb{R}^n \setminus B(x,r)} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy \leqslant c r^{\alpha-n/p}\|f\|_p.$$

Lemma 2.39 implies

$$\begin{aligned}
|I_\alpha f(x)| \leqslant \int_{\mathbb{R}^n} \frac{|f(y)|}{|x-y|^{n-\alpha}}\,dy &= \int_{B(x,r)} \dots\,dy + \int_{\mathbb{R}^n \setminus B(x,r)} \dots\,dy \\
&\leqslant c\left(r^{\alpha} Mf(x) + r^{\alpha-(n/p)}\|f\|_p\right).
\end{aligned}$$

By choosing

$$r = \left(\frac{Mf(x)}{\|f\|_p} \right)^{-p/n},$$

we obtain

$$|I_\alpha f(x)| \leqslant cMf(x)^{1-\alpha p/n} \|f\|_p^{\alpha p/n}.$$

By raising both sides to the power $p^* = np/(n - \alpha p)$, we have

$$|I_\alpha f(x)|^{p^*} \leqslant cMf(x)^p \|f\|_p^{(\alpha p/n)p^*}$$

The Hardy-Littlewood theorem II (Theorem 2.22) implies

$$\int_{\mathbb{R}^n} |I_\alpha f(x)|^{p^*} \, dy \leqslant c\|f\|_p^{(\alpha p/n)p^*} \int_{\mathbb{R}^n} (Mf(x))^p \, dx \leqslant c\|f\|_p^{(\alpha p/n)p^*} \|f\|_p^p$$

and thus

$$\|I_\alpha f\|_{p^*} \leqslant c\|f\|_p^{\alpha p/n + p/p^*} = c\|f\|_p. \qquad \square$$

Corollary 2.41 (The Sobolev inequality). If $1 < p < n$, there exists a constant $c = c(n,p) > 0$ such that

$$\|u\|_{p^*} \leqslant c\|\nabla u\|_p$$

for every $u \in C_0^1(\mathbb{R}^n)$.

Proof. Lemma 2.38

$$|u(x)| \leqslant \frac{1}{\omega_{n-1}} I_1(\nabla u)(x) \quad \text{for every} \quad x \in \mathbb{R}^n,$$

Thus Theorem 2.40 implies

$$\|u\|_{p^*} \leqslant c\|I_1(\nabla u)\|_{p^*} \leqslant c\|\nabla u\|_p. \qquad \square$$

Remark 2.42. Let $\Omega \subset \mathbb{R}^n$ be an open set and $u \in C_0^1(\Omega)$. By defining $u(x) = 0$ for every $x \in \mathbb{R}^n \setminus \Omega$, we have

$$\left(\int_\Omega |u|^{p^*} \, dx \right)^{1/p^*} \leqslant c \left(\int_\Omega |\nabla u|^p \, dx \right)^{1/p}.$$

In this section we consider the definition and properties of convolution. Convolutions are used to approximate and mollify L^p functions. Moreover, many operators in harmonic analysis and partial differential equations can be written as a convolution. Approximations of the identity converge in L^p and pointwise almost everywhere under appropriate assumptions. As an application we show that $C_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$. Solution to the Dirichlet problem with L^p boundary values for the Laplace equation in the upper half space can be expressed as a convolution against the Poisson kernel.

3 Convolutions

In this section we work with the Lebesgue measure on $\mathbb{R}^n$.

3.1 Two additional properties of the L^p spaces

First we consider the question of approximation of L^p functions by compactly supported continuous functions.

Definition 3.1. The support of a function $f : \mathbb{R}^n \to [-\infty, \infty]$ is

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}.$$

If $f \in C(\mathbb{R}^n)$ and $\operatorname{supp} f$ is a compact set, then we denote $f \in C_0(\mathbb{R}^n)$ and say that f is a compactly supported continuous function.

THE MORAL: A function is compactly supported if and only if it is zero in the complement of a sufficiently large ball. Thus a compactly supported function is identically zero far way from the origin.

Remark 3.2. $C_0(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ for every $1 \leqslant p \leqslant \infty$. Thus compactly supported continuous functions belong to every $L^p(\mathbb{R}^n)$ with $1 \leqslant p \leqslant \infty$.

Reason. $\boxed{1 \leqslant p < \infty}$

$$\int_{\mathbb{R}^n} |f(x)|^p \, dx = \int_{\operatorname{supp} f} |f(x)|^p \, dx \leqslant \sup_{x \in \operatorname{supp} f} |f(x)|^p |\operatorname{supp} f| < \infty,$$

since a continuous function assumes its maximum in a compact set and a compact set has finite Lebesgue measure.

$\boxed{p=\infty}$

$$|f(x)| \leqslant \sup_{x \in \operatorname{supp} f} |f(x)| |\operatorname{supp} f| < \infty,$$

from which it follows that $\|f\|_\infty < \infty$. ■

Theorem 3.3. Assume that $1 \leqslant p < \infty$. Then $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$.

THE MORAL: This means that for every $\varepsilon > 0$ there is a function $g \in C_0(\mathbb{R}^n)$ such that $\|f - g\|_p < \varepsilon$. Equivalently, any function $f \in L^p(\mathbb{R}^n)$ can be approximated by functions $f_i \in C_0(\mathbb{R}^n)$ in $L^p(\mathbb{R}^n)$, that is, $\|f_i - f\|_p \to 0$ as $i \to \infty$.

WARNING: $C_0(\mathbb{R}^n)$ is not dense in $L^\infty(\mathbb{R}^n)$, because the limit of continuous functions in $L^\infty(\mathbb{R}^n)$ is a continuous function. If this would be true, then this would imply that all functions $L^\infty(\mathbb{R}^n)$ are continuous, which is not the case. There is also another reson why this is not true. The constant function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = 1$ cannot be approximated by compactly supported functions in $L^\infty(\mathbb{R})$.

Proof. Assume $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p < \infty$.

$\boxed{(1)}$ Let $f_i = f\chi_{B(0,i)}$, $i = 1, 2, \ldots$. Then

$$\lim_{i\to\infty} f_i(x) = f(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

By the Lebesgue dominated convergence theorem

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} |f_i - f|^p \, dx = \int_{\mathbb{R}^n} \lim_{i\to\infty} |f_i - f|^p \, dx = 0.$$

Observe that

$$|f_i(x) - f(x)|^p \leqslant (|f_i(x)| + |f(x)|)^p \leqslant 2^p |f|^p \in L^1(\mathbb{R}^n))$$

gives an integrable majorant.

Thus compactly supported functions in $L^p(\mathbb{R}^n)$ are dense in $L^p(\mathbb{R}^n)$ and we may assume that f is such a function.

$\boxed{(2)}$ Since $f = f_+ - f_-$, we may assume that $f \geqslant 0$ and that $f = 0$ outside a compact set. Indeed this set can be chosen to be a closed ball $\overline{B(0,i)}$ for i large enough.

$\boxed{(3)}$: Since $f \geqslant 0$ is measurable, there exists an increasing sequence of simple functions g_i such that

$$\lim_{i\to\infty} g_i(x) = f(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

Since $0 \leqslant g_i \leqslant f$, we have

$$\int_{\mathbb{R}^n} g_i^p \, dx \leqslant \int_{R^n} f^p \, dx < \infty$$

and thus $g_i \in L^p(\mathbb{R}^n)$, $i = 1, 2, \ldots$. By the Lebesgue dominated convergence theorem

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} |g_i - f|^p \, dx = \int_{\mathbb{R}^n} \lim_{i\to\infty} |g_i - f|^p \, dx = 0.$$

Observe that

$$|g_i(x) - f(x)|^p \leqslant (|g_i(x)| + |f(x)|)^p \leqslant 2^p |f|^p \in L^1(\mathbb{R}^n))$$

gives an integrable majorant.

Thus we may assume that f is a nonnegative simple function with a compact support.

(4) A simple function can be represented as the finite sum $f = \sum_{i=1}^{k} a_i \chi_{A_i}$, where the sets A_i are bounded, measurable and pairwise disjoint, $a_i \geqslant 0$. Thus we may assume that $f = \chi_A$, where A is a bounded measurable set.

(5) By the approximation result for measurable sets, there exists an open set $G \supset A$ and a closed set $F \subset A$ such that $|G \setminus F|^{1/p} < \varepsilon$, where $\varepsilon > 0$. The set F is compact, since it is closed and bounded.

(6) We recall the following version of the Urysohn lemma. Assume that $G \subset \mathbb{R}^n$ is an open set and that $F \subset G$ a compact set. Then there exists a continuous function $g : \mathbb{R}^n \to \mathbb{R}$ such that

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

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

(3) $\operatorname{supp} g$ is a compact subset of G.

Reason. Let

$$U = \{x \in \mathbb{R}^n : \operatorname{dist}(x, F) < \frac{1}{2} \operatorname{dist}(F, \mathbb{R}^n \setminus G)\}.$$

Then $F \subset U \subset \overline{U} \subset G$, U is open and $\overline{U}$ is compact. Define $g : \mathbb{R}^n \to \mathbb{R}$,

$$g(x) = \frac{\operatorname{dist}(x, \mathbb{R}^n \setminus U)}{\operatorname{dist}(x, F) + \operatorname{dist}(x, \mathbb{R}^n \setminus U)}, \tag{3.1}$$

where $\operatorname{dist}(x, A) = \inf\{|x - y| : y \in A\}$ is the distance of x from A.

It is clear that $0 \leqslant g(x) \leqslant 1$ for every $x \in \mathbb{R}^n$.

Let $x \in F$. Then $\operatorname{dist}(x, F) = 0$. Since $F \subset U$, there exists $r > 0$ such that $B(x, r) \subset U$. This implies $\operatorname{dist}(x, \mathbb{R}^n \setminus U) \geqslant r > 0$ and thus $g(x) = 1$.

Moreover, $\operatorname{supp} g = \overline{\{x \in \mathbb{R}^n : g(x) \neq 0\}} \subset \overline{U}$, which is a closed and bounded set and thus compact.

We claim that $x \mapsto \operatorname{dist}(x, A)$ is continuous for every $A \neq \emptyset$. Let $x, x' \in \mathbb{R}^n$. Then

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

for every $y \in A$. This implies $\operatorname{dist}(x, A) - |x - x'| \leqslant \operatorname{dist}(x', A)$ and thus $\operatorname{dist}(x, A) - \operatorname{dist}(x', A) \leqslant |x - x'|$. By switching the roles of x and x' we have $\operatorname{dist}(x', A) - \operatorname{dist}(x, A) \leqslant |x - x'|$, from which we conclude

$$|\operatorname{dist}(x, A) - \operatorname{dist}(x', A)| \leqslant |x - x'|.$$

Note that $\operatorname{dist}(x, A)$ is even Lipschitz continuous with the constant 1. This implies that g is continuous. Thus g has all the required properties. ■

THE MORAL: This shows that there exists a continuous function g which satisfies $\chi_F \leqslant f \leqslant \chi_G$. Note that it is easy to find semicontinuous functions with this property, since χ_F and χ_G will do.

(7) Let g be a function as in (6). Note that $\chi_A(x) - g(x) = 1 - 1 = 0$ for every $x \in F$, $\chi_A(x) - g(x) = 0 - 0 = 0$ for every $x \in \mathbb{R}^n \setminus G$ and $|\chi_A - g| \leqslant 1$. Thus

$$\|f - g\|_p = \left(\int_{\mathbb{R}^n} |\chi_A - g|^p \, dx \right)^{1/p} \leqslant |G \setminus F|^{1/p} < \varepsilon.$$

□

Remark 3.4. The proof above shows that simple functions are dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p \leqslant \infty$.

Remark 3.5. Let us briefly discuss the question of separability of the L^p spaces. Recall that a metric space is separable, if there is a countable dense subset of the space. For the sequence spaces l^p with $1 \leqslant p < \infty$ this is an exercise. However, the space l^∞ is not separable,

Reason. Let S be the collection of all sequences consisting of ones and zeros. If x and y are distinct, then $\|x - y\|_\infty \geqslant 1$. Since S is an uncountable subset of l^∞ and every pair of points in S is at least a unit distance apart, there can be no countable dense subset of l^∞. ■

Generally, we can expect similar assertions for the L^p spaces, Indeed, the space $L^\infty(\mathbb{R}^n)$ is not separable and $L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$ are separable. For example, $L^1([0,1])$ is separable, since the collection of rational linear combinations of the characteristic functions of those sets that are finite unions of intervals with rational endpoints gives a countable dense subset. However, for other measures that the Lebesgue measure separability depends on the measure.

Then we study a useful continuity property of the integral. This will be an important tool in proving that convolution approximations converge to the original function. Moreover, it can be used to prove the Riemann-Lebesgue lemma, which asserts that the Fourier transform $\widehat{f}(\xi)$ of a function $f \in L^1(\mathbb{R}^n)$ has the property $\lim_{|\xi| \to \infty} \widehat{f}(\xi) = 0$ (exercise).

Theorem 3.6. Assume $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p < \infty$. Then

$$\lim_{y \to 0} \int_{\mathbb{R}^n} |f(x+y) - f(x)|^p \, dx = 0.$$

THE MORAL: Let $\tau_y f(x) = f(x+y)$, $y \in \mathbb{R}^n$, be the translation. Then

$$\lim_{y \to 0} \|\tau_y f - f\|_p \to 0.$$

Thus the translation $\tau_y f$ depends continuously on y at $y = 0$.

WARNING: The claim does not hold when $p = \infty$. In fact, if $f \in L^\infty(\mathbb{R}^n)$ satisfies $\lim_{y \to 0} \|\tau_y f - f\|_\infty = 0$, then f can be redefined on a set of measure zero so that it becomes uniformly continuous (exercise).

Reason. Let $f:\mathbb{R}\to\mathbb{R}$, $f=\chi_{[0,\infty)}$. Then

$$\operatorname*{ess\,sup}_{x\in\mathbb{R}} |f(x+y)-f(x)| = 1 \quad \text{for every} \quad y\neq 0.$$

■

Proof. Let $\varepsilon>0$ and $y\in\mathbb{R}^n$. By Theorem 3.3, there exists $g\in C_0(\mathbb{R}^n)$ such that

$$\left(\int_{\mathbb{R}^n} |f(x)-g(x)|^p\,dx\right)^{1/p} < \frac{\varepsilon}{3}.$$

The translation invariance of the Lebesgue integral implies

$$\left(\int_{\mathbb{R}^n} |f(x+y)-g(x+y)|^p\,dx\right)^{1/p} = \left(\int_{\mathbb{R}^n} |f(x)-g(x)|^p\,dx\right)^{1/p}.$$

Since $g\in C_0(\mathbb{R}^n)$, there exists $r>0$ such that $g(x)=0$ for every $x\in\mathbb{R}^n\setminus B(0,r)$ and thus g is uniformly continuous in $\mathbb{R}^n$. Here we used the fact that a continuous function is uniformly continuous on compact sets. Therfore, there exists $0<\delta\leqslant 1$ such that

$$|g(x+y)-g(x)| < \frac{\varepsilon}{3|B(0,r+1)|^{1/p}} \quad \text{for every} \quad x\in\mathbb{R}^n, \quad |y|<\delta.$$

Since $g(x+y)-g(x)=0$ for every $x\in\mathbb{R}^n\setminus B(0,r+1)$, we have

$$\begin{aligned}\left(\int_{\mathbb{R}^n} |g(x+y)-g(x)|^p\,dx\right)^{1/p} &= \left(\int_{B(0,r+1)} |g(x+y)-g(x)|^p\,dx\right)^{1/p}\\ &< \frac{\varepsilon}{3|B(0,r+1)|^{1/p}}|B(0,r+1)|^{1/p} = \frac{\varepsilon}{3}.\end{aligned}$$

Therefore

$$\begin{aligned}&\left(\int_{\mathbb{R}^n} |f(x+y)-f(x)|^p\,dx\right)^{1/p}\\ &\quad\leqslant \left(\int_{\mathbb{R}^n} |f(x+y)-g(x+y)|^p\,dx\right)^{1/p} + \left(\int_{\mathbb{R}^n} |g(x+y)-g(x)|^p\,dx\right)^{1/p}\\ &\qquad + \left(\int_{\mathbb{R}^n} |g(x)-f(x)|^p\,dx\right)^{1/p}\\ &\quad< \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

□

3.2 Definition of the convolution

We begin with a formal definition of the convolution.

Definition 3.7. Assume that $f,g:\mathbb{R}^n\to[-\infty,\infty]$ are measurable functions. On a formal level, the convolution $f*g:\mathbb{R}^n\to[-\infty,\infty]$ is defined by

$$(f*g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)\,dy,$$

when this makes sense.

W A R N I N G : It is not clear that the integrand is a measurable function and that the integral is well defined. This requires further analysis.

Remark 3.8. The function $y \mapsto f(y)g(x-y)$ is a measurable function for a fixed $x \in \mathbb{R}^n$.

Reason. Let $U \subset \mathbb{R}$ be an open set. The translation function $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $\Phi(y) = x - y$, maps measurable sets to measurable sets, so that

$$(g \circ \Phi)^{-1}(U) = \Phi^{-1}(g^{-1}(U))$$

is a measurable set. This shows that $y \mapsto (g \circ \Phi)(y) = g(x-y)$ is a measurable function. Thus $y \mapsto f(y)g(x-y)$ is a measurable function as a product of two measurable functions. ■

T H E M O R A L : The convolution is well defined for nonnegative functions f and g, but the integral may be infinite for every $x \in \mathbb{R}^n$.

A more careful analysis is needed to deal with sign changing functions. Then we need conditions under which the integrals of the positive and negative parts are finite. We begin with considering the measurability question with respect to the product Lebesgue measure in $\mathbb{R}^n \times \mathbb{R}^n$. This is needed in the application of Fubini's theorem, which ensures almost everywhere finiteness of $|f| * |g|$ under appropriate conditions.

Remarks 3.9:

(1) Assume that $f : \mathbb{R}^n \to [-\infty, \infty]$ is a measurable function. There exists a Borel function $\phi : \mathbb{R}^n \to [0, \infty]$ such that $\phi = f$ almost everywhere.

Reason. Recall that ϕ is Borel function, if the preimage $\phi^{-1}(U) = \{x \in \mathbb{R}^n : \phi(x) \in U\}$ is a Borel set, when $U \subset \mathbb{R}$ is a open set.
Assume first that f is nonnegative. Since f is a nonnegative measurable function, there exists an increasing sequence of simple functions f_i, $i = 1, 2, \dots$, such that

$$f(x) = \lim_{i \to \infty} f_i(x)$$

for every $x \in \mathbb{R}^n$. Each simple function f_i can be written as a finite sum

$$f_i = \sum_j a_j \chi_{A_j},$$

where $A_j = f^{-1}(\{a_j\})$ are measurable sets. For each measurable set A_j there is a Borel set B_i such that $|A_i \setminus B_i| = 0$. Then $\phi_i = \sum_j a_j \chi_{B_j}$ is a Borel function, $0 \leqslant \phi_i \leqslant f_i$ and $\phi_i = f_i$ almost everywhere. Note that the sequence ϕ_i is not necessary increasing and the limit $\lim_{i\to\infty} \phi_i$ does not necessarily exist, but

$$\phi = \liminf_{i \to \infty} \phi_i$$

is a Borel function and $\phi = f$ almost everywhere.

In the general case $f : \mathbb{R}^n \to [-\infty,\infty]$, we write $f = f^+ - f^-$ and consider the positive and negative parts separately. ∎

Note that the integrals do not change if we replace f with ϕ. In particular, we may assume that f and g are Borel functions in the definition of convolution.

(2) If $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$ are Borel functions, then $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $F(x, y) = f(y)g(x-y)$ is a Borel function.

Reason. The functions

$$\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,\ \Phi(x,y) = x - y \quad \text{and} \quad \Psi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n,\ \Psi(x,y) = y,$$

are continuous and $F(x,y) = f(\Psi(x,y))g(\Phi(x,y))$, that is, $F = (f\circ\Psi)\cdot(g\circ\Phi)$. Let $U \subset \mathbb{R}$ be an open set. Since f is a Borel function, the preimage $f^{-1}(U)$ is a Borel set in $\mathbb{R}^n$. Since Ψ is a continuous function and the preimage of a Borel set is a Borel set in a continuous function, we conclude that

$$(f\circ\Psi)^{-1}(U) = \Psi^{\cdot -1}(f^{-1}(U))$$

is a Borel set in $\mathbb{R}^{2n}$. In same way, we see that $g\circ\Phi$ is a Borel function and thus F is a Borel function as a product of two Borel functions. Since the product Lebesgue measure on $\mathbb{R}^n \times \mathbb{R}^n$ is a Borel measure, we conclude that F is measurable. ∎

THE MORAL: We may assume that the functions f and g in the definition of convolution are Borel functions. In this case the Fubini's theorem is available as a tool to show that the integral is well defined.

The next result settles the integrability questions in the definition of the convolution under certain assumptions.

Theorem 3.10 (Young's theorem). Assume that $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p \leqslant \infty$ and $g \in L^1(\mathbb{R}^n)$. Then $(f*g)(x)$ exists for almost every $x \in \mathbb{R}^n$ and $\|f*g\|_p \leqslant \|f\|_p\|g\|_1$.

THE MORAL: The convolution of an L^p function and L^1 function is well defined. Moreover, it is an L^p function.

WARNING: $f, g \in L^1(\mathbb{R}^n)$ does not imply that the function $y \mapsto f(y)g(x-y)$ is in $L^1(\mathbb{R}^n)$ for a fixed $x \in \mathbb{R}^n$. A product of integrable functions is not necessarily integrable. However, $\|f*g\|_1 \leqslant \|f\|_1\|g\|_1$ and thus $f*g \in L^1(\mathbb{R}^n)$.

Proof. $\boxed{p=1}$ First assume that f and g are nonnegative. Then $f(y)g(x-y)$ is a nonnegative measurable function on $\mathbb{R}^{2n}$ and by Fubini's theorem for nonnegative functions

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(y)g(x-y)\,dy\,dx = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(y)g(x-y)\,dx\,dy.$$

On the other hand,

$$\begin{aligned}\int_{\mathbb{R}^n}(f*g)(x)\,dx &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(y)g(x-y)\,dy\,dx\\ &= \int_{\mathbb{R}^n} f(y)\left(\int_{\mathbb{R}^n} g(x-y)\,dx\right)dy\\ &= \int_{\mathbb{R}^n} f(y)\left(\int_{\mathbb{R}^n} g(x)\,dx\right)dy\\ &= \int_{\mathbb{R}^n} f(y)\,dy\int_{\mathbb{R}^n} g(x)\,dx\end{aligned}$$

Thus we have $\|f*g\|_1 = \|f\|_1\|g\|_1$ and the claim holds in this case.

Let us then consider the general case. By the begining of the proof $|f|*|g|$ exists almost everywhere. Thus for almost every x the function $y \mapsto |f(y)g(x-y)|$ is integrable. This means that for almost every x the function $y \mapsto f(y)g(x-y)$ is integrable and we conclude that $f*g$ exists almost everywhere. Since $|f*g| \leqslant |f|*|g|$, we have

$$\|f*g\|_1 \leqslant \||f|*|g|\|_1 = \|f\|_1\|g\|_1.$$

$\boxed{p=\infty}$

$$\begin{aligned}|(f*g)(x)| &\leqslant \int_{\mathbb{R}^n}|f(y)||g(x-y)|\,dy\\ &\leqslant \operatorname*{ess\,sup}_{y\in\mathbb{R}^n}|f(y)|\int_{\mathbb{R}^n}|g(x-y)|\,dy = \|f\|_\infty\|g\|_1.\end{aligned}$$

This implies that $\|f*g\|_\infty \leqslant \|f\|_\infty\|g\|_1$.

$\boxed{1<p<\infty}$ By Hölder's inequality

$$\begin{aligned}|(f*g)(x)| &\leqslant \int_{\mathbb{R}^n}|f(y)||g(x-y)|\,dy\\ &= \int_{\mathbb{R}^n}|f(y)||g(x-y)|^{1/p}|g(x-y)|^{1/p'}\,dy\\ &\leqslant \left(\int_{\mathbb{R}^n}|f(y)|^p|g(x-y)|\,dy\right)^{1/p}\left(\int_{\mathbb{R}^n}|g(x-y)|\,dy\right)^{1/p'}.\end{aligned}$$

This implies

$$|(f*g)(x)|^p \leqslant \|g\|_1^{p/p'}\int_{\mathbb{R}^n}|f(y)|^p|g(x-y)|\,dy$$

and thus

$$\begin{aligned}\int_{\mathbb{R}^n}|(f*g)(x)|^p\,dx &\leqslant \|g\|_1^{p/p'}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}|f(y)|^p|g(x-y)|\,dy\,dx\\ &= \|g\|_1^{p/p'}\int_{\mathbb{R}^n}|f(y)|^p\left(\int_{\mathbb{R}^n}|g(x-y)|\,dx\right)dy \qquad \text{(Fubini)}\\ &= \|g\|_1^{p/p'}\|g\|_1\|f\|_p^p = \|g\|_1^p\|f\|_p^p.\end{aligned}$$

□

Remark 3.11. Let $f \in L^1(\mathbb{R}^n)\setminus L^2(\mathbb{R}^n)$ such that $f(x) = f(-x)$. By Young's inequality $(f * f)(x) < \infty$ for almost every $x \in \mathbb{R}^n$. However,

$$(f * f)(0) = \int_{\mathbb{R}^n} f(y)f(-y)\,dy = \int_{\mathbb{R}^n} |f(y)|^2\,dy = \infty,$$

which shows that $f * f$ blows up at $x = 0$.

Theorem 3.12. Assume that $1 \leqslant p \leqslant \infty$, $f \in L^p(\mathbb{R}^n)$ and $g \in L^{p'}(\mathbb{R}^n)$. Then $(f * g)(x)$ exists for every $x \in \mathbb{R}^n$ and $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_{p'}$. Moreover, the function $f * g$ is uniformly continuous in $\mathbb{R}^n$.

THE MORAL: The convolution of an L^p function and $L^{p'}$ function is well defined. Moreover, it is a bounded and continuous function.

Proof. In the general case, either p or p' is finite (or both). Assume $p < \infty$. By Hölder's inequality

$$\begin{aligned}
|(f * g)(x)| &= \left| \int_{\mathbb{R}^n} f(x-y)g(y)\,dy \right| \\
&\leqslant \left(\int_{\mathbb{R}^n} |f(x-y)|^p\,dy \right)^{1/p} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{1/p'} \\
&= \|f\|_p \|g\|_{p'} < \infty
\end{aligned}$$

for every $x \in \mathbb{R}^n$. This implies $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_{p'}$.

Moreover, by Hölder's inequality and the translation invariance of the Lebesgue integral, we have

$$\begin{aligned}
|(f * g)(x) - (f * g)(z)| &= \left| \int_{\mathbb{R}^n} (f(x-y) - f(z-y))g(y)\,dy \right| \\
&\leqslant \left(\int_{\mathbb{R}^n} |f(x-y) - f(z-y)|^p\,dy \right)^{1/p} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{1/p'} \\
&= \|\tau_{-x}f - \tau_{-z}f\|_p \|g\|_{p'} = \|\tau_{z-x}f - f\|_p \|g\|_{p'}.
\end{aligned}$$

Theorem 3.6 implies that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|\tau_{z-x}f - f\|_p < \varepsilon$ whenever $|z - x| < \delta$. Thus

$$|(f * g)(x) - (f * g)(z)| < \varepsilon \|g\|_{p'}$$

for every $x \in \mathbb{R}^n$ and $|z| < \delta$. Therefore, $f * g$ is uniformly continuous. □

The following lemma shows that the convolution regarded as a multiplication in $L^1(\mathbb{R}^n)$ satisfies certain standard algebraic laws.

Lemma 3.13. Assume $f, g, h \in L^1(\mathbb{R}^n)$ and $a, b \in \mathbb{R}$. Then the following claims are true:

(1) (Commutative law) $f * g = g * f$.

(2) (Associative law)$f * (g * h) = (f * g) * h$.

(3) (Distributive law) $(af + bg) * h = a(f * h) + b(g * h)$.

Proof. (Exercise) □

3.3 Approximations of the identity

The previous lemma, Riesz-Fischer theorem and Young's theorem show that $L^1(\mathbb{R}^n)$ is a commutative Banach algebra with the convolution as a product. This algebra does not have a multiplicative identity, that is, there does not exist $\phi \in L^1(\mathbb{R}^n)$ such that $\phi * f = f$ for every $f \in L^1(\mathbb{R}^n)$.

Reason. Assume that there exists such a ϕ. Then, in particular, $\phi * f = f$ for every $f \in L^\infty(\mathbb{R}^n)$ with a compact support. Theorem 3.12 implies that $\phi * f$ is continuous. Since $\phi * f = f$, this shows that every $f \in L^\infty(\mathbb{R}^n)$ with a compact support is continuous. This is not true, take $f = \chi_{B(0,1)}$, for example. ■

However, there exists approximations of the identity in the sense that there exists a collection of functions $\phi_\varepsilon \in L^1(\mathbb{R}^n)$ such that $\phi_\varepsilon * f \to f$ in $L^1(\mathbb{R}^n)$ as $\varepsilon \to 0$. In fact, the limit exists in $L^p(\mathbb{R}^n)$ and pointwise under appropriate assumptions. This gives a very useful method to produce approximations of functions in $L^p(\mathbb{R}^n)$.

Definition 3.14. Let $\phi \in L^1(\mathbb{R}^n)$, $\varepsilon > 0$. Define

$$\phi_\varepsilon(x) = \frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right), \quad x \in \mathbb{R}^n.$$

Such a collection of functions is called an approximation of the identity.

THE MORAL: Smaller values of $\varepsilon > 0$ produce higher peaks and smaller supports.

Examples 3.15:

(1) Let $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\phi(x) = \frac{\chi_{B(0,1)}(x)}{|B(0,1)|}.$$

Then

$$\phi_\varepsilon(x) = \frac{1}{\varepsilon^n}\frac{\chi_{B(0,1)}(x/\varepsilon)}{|B(0,1)|} = \frac{\chi_{B(0,\varepsilon)}(x)}{|B(0,\varepsilon)|}, \quad \varepsilon > 0.$$

Assume $f \in L^1(\mathbb{R}^n)$. Then

$$(f * \phi_\varepsilon)(x) = \int_{\mathbb{R}^n} f(y)\phi_\varepsilon(x-y)\,dy = \frac{1}{|B(x,\varepsilon)|}\int_{B(x,\varepsilon)} f(y)\,dy$$

is the integral average of over the ball $B(x,\varepsilon)$. By Young's inequality

$$\|f * \phi_\varepsilon\|_1 \leqslant \|f\|_1\|\phi_\varepsilon\|_1 = \|f\|_1 \quad \text{for every} \quad \varepsilon > 0,$$

since $\|\phi_\varepsilon\|_1 = 1$ for every $\varepsilon > 0$.

(2) Let $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi(x) = \begin{cases} \exp\left(\frac{1}{|x|^2-1}\right), & x \in B(0,1), \\ 0, & x \in \mathbb{R}^n \setminus B(0,1). \end{cases}$$

Then $\varphi \in C_0(\mathbb{R}^n)$ and thus $\varphi \in L^1(\mathbb{R}^n)$ with $0 < \|\varphi\|_1 < \infty$. Let

$$\phi(x) = \frac{\varphi(x)}{\|\varphi\|_1}, \quad x \in \mathbb{R}^n.$$

Then $\phi_\varepsilon \in C_0(\mathbb{R}^n)$ and $\operatorname{supp}\phi_\varepsilon = \overline{B(0,\varepsilon)}$. Moreover

$$\begin{aligned}
\int_{\mathbb{R}^n} \phi_\varepsilon(x)\,dx &= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} \phi\left(\frac{x}{\varepsilon}\right) dx \\
&= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} \phi(y)\varepsilon^n\,dy \qquad (y = \frac{x}{\varepsilon}, dx = \varepsilon^n\,dy) \\
&= \int_{\mathbb{R}^n} \phi(y)\,dy \\
&= \int_{\mathbb{R}^n} \frac{\varphi(x)}{\|\varphi\|_1}\,dx = \frac{\|\phi\|_1}{\|\varphi\|_1} = 1 \quad \text{for every} \quad \varepsilon > 0.
\end{aligned}$$

Young's inequality implies

$$\|f * \phi_\varepsilon\|_1 \leqslant \|f\|_1\|\phi_\varepsilon\|_1 = \|f\|_1 \quad \text{for every} \quad \varepsilon > 0.$$

The function ϕ_ε is called the standard mollifier.

Lemma 3.16. Let $\phi \in L^1(\mathbb{R}^n)$.

(1) $\displaystyle\int_{\mathbb{R}^n} \phi_\varepsilon(x)\,dx = \int_{\mathbb{R}^n} \phi(x)\,dx$ for every $\varepsilon > 0$.

(2) $\displaystyle\lim_{\varepsilon\to 0}\int_{\mathbb{R}^n\setminus B(0,r)} |\phi_\varepsilon(x)|\,dx = 0$ for every $r > 0$.

THE MORAL: The condition (1) explains the scaling factors in the definition of ϕ_ε. These are chosen so that the L^1 norm of ϕ_ε is independent of $\varepsilon > 0$. The condition (2) tells that the function ϕ_ε is concentrated in a small neighbourhood of the point x.

Remark 3.17. The condition (2) is clear, if the support of ϕ is a compact set.

Proof. (1) A change of variables.

(2)

$$\begin{aligned}
\int_{\mathbb{R}^n\setminus B(0,r)} |\phi_\varepsilon(x)|\,dx &= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n\setminus B(0,r)} \left|\phi\left(\frac{x}{\varepsilon}\right)\right| dx \\
&= \int_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})} |\phi(y)|\,dy \qquad (y = \frac{x}{\varepsilon} \Rightarrow dx = \varepsilon^n\,dy) \\
&= \int_{\mathbb{R}^n} |\phi(x)|\chi_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})}(x)\,dx \to 0
\end{aligned}$$

by the Lebesgue dominated convergence theorem with the integrable majorant

$$|\phi|\chi_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})} \leqslant |\phi| \in L^1(\mathbb{R}^n)) \quad \text{for every} \quad \varepsilon > 0.$$

□

Remarks 3.18:

(1) $\|\phi_\varepsilon\|_p = \left(\frac{1}{\varepsilon^n}\right)^{(p-1)/p}\|\phi\|_p$, $1 \leqslant p < \infty$. (Exercise)

(2) $\|\phi_\varepsilon\|_\infty = \frac{1}{\varepsilon^n}\|\phi\|_\infty$. (Exercise)

3.4 Pointwise convergence

There is a connection between the approximations of the identity and the Hardy-Littlewood maximal function. Recall that a function $\phi : \mathbb{R}^n \to \mathbb{R}$ is radial, if its value only depends on $|x|$. Thus a nonnegative radial function is of the form $f(x) = \varphi(|x|)$ for some function $\varphi : \mathbb{R}^+ \to \mathbb{R}^+$. We say that a radial function is decreasing, if $|x| \geqslant |y|$ implies $\phi(x) \leqslant \phi(y)$.

Theorem 3.19. Assume that $\phi \in L^1(\mathbb{R}^n)$ is nonnegative, radial and decreasing Then

$$\sup_{\varepsilon>0} |(\phi_\varepsilon * f)(x)| \leqslant \|\phi\|_1 Mf(x)$$

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

THE MORAL: The Hardy-Littlewood maximal operator gives a majorant to many other operators as well. In this sense the maximal function controls the averages of a function with respect to any radial and decreasing function.

Proof. First assume in addition to the given hypotheses that ϕ is a simple function in the form

$$\phi(x) = \sum_i a_i \chi_{B(0,r_i)}$$

with $a_i > 0$. The sum here is over finitely many indices only. Then

$$\|\phi\|_1 = \sum_i a_i |B(0,r_i)|.$$

Moreover,

$$\begin{aligned}
|(\phi_\varepsilon * f)(x)| &= \left| \int_{\mathbb{R}^n} f(x-y)\phi_\varepsilon(y)\,dy \right| = \left| \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} f(x-y)\phi\left(\frac{y}{\varepsilon}\right) dy \right| \\
&= \left| \int_{\mathbb{R}^n} f(x-\varepsilon z)\phi(z)\,dz \right| \qquad \left(z = \frac{y}{\varepsilon},\ y = \varepsilon z,\ dy = \varepsilon^n\,dz\right) \\
&= \left| \sum_i \int_{B(0,r_i)} f(x-\varepsilon z) a_i\,dz \right| \leqslant \sum_i a_i \int_{B(0,r_i)} |f(x-\varepsilon z)|\,dz \\
&= \sum_i a_i |B(0,r_i)| \frac{1}{|B(0,r_i)|} \int_{B(0,r_i)} |f(x-\varepsilon z)|\,dz,
\end{aligned}$$

where

$$\begin{aligned}
\frac{1}{|B(0,r_i)|} \int_{B(0,r_i)} |f(x-\varepsilon z)|\,dz &= \frac{1}{\varepsilon^n |B(0,r_i)|} \int_{B(x,\varepsilon r_i)} |f(y)|\,dy \\
&\qquad (y = x-\varepsilon z,\ z = (y-x)/\varepsilon,\ dz = \varepsilon^{-n}\,dy) \\
&= \frac{1}{|B(x,\varepsilon r_i)|} \int_{B(x,\varepsilon r_i)} |f(y)|\,dy \leqslant Mf(x).
\end{aligned}$$

Thus

$$|(\phi_\varepsilon * f)(x)| \leqslant \sum_i a_i |B(0,r_i)| Mf(x) = \|\phi\|_1 Mf(x).$$

Then we consider the general case. Since ϕ is nonnegative, radial and decreasing, there is an increasing sequence of simple functions ϕ_j such that $\phi_j(x) \to \phi(x)$ for every $x \in \mathbb{R}^n$ as $j \to \infty$. By the Lebesgue monotone convergence theorem

$$\begin{aligned}|(\phi_\varepsilon * f)(x)| &\leqslant \int_{\mathbb{R}^n} |f(x-y)|\phi_\varepsilon(y)\,dy \\ &= \int_{\mathbb{R}^n} |f(x-y)| \lim_{j\to\infty} (\phi_j)_\varepsilon(y)\,dy \\ &= \lim_{j\to\infty} \int_{\mathbb{R}^n} |f(x-y)|(\phi_j)_\varepsilon(y)\,dy \\ &\leqslant \lim_{j\to\infty} \|\phi_j\|_1 Mf(x) = \|\phi\|_1 Mf(x)\end{aligned}$$

for every $x \in \mathbb{R}^n$. □

Theorem 3.20. Assume that $\phi \in L^1(\mathbb{R}^n)$ is nonnegative, radial and decreasing, $a = \int_{\mathbb{R}^n} \phi(x)\,dx$ and $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p \leqslant \infty$. Then

$$\lim_{\varepsilon\to 0} (\phi_\varepsilon * f)(x) = af(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

T H E M O R A L : This is the Lebesgue differentiation theorem for approximations of the identity. This shows that the convolution approximations can be seen as weighted averages of the function.

Proof. $\boxed{1 \leqslant p < \infty}$ Define a maximal operator related to the aproximation of the unity by

$$M_\phi f(x) = \sup_{\varepsilon>0} |(\phi_\varepsilon * f)(x)| \leqslant \|\phi\|_1 Mf(x).$$

Then Theorem 3.19 gives

$$M_\phi f(x) \leqslant \|\phi\|_1 Mf(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

By the weak type estimate for the Hardy-Littlewood maximal function, see Theorem 2.18, for $f \in L^1(\mathbb{R}^n)$, we have

$$|\{x \in \mathbb{R}^n : M_\phi f(x) > \lambda\}| \leqslant \frac{c}{\lambda}\|f\|_1 \quad \text{for every} \quad \lambda > 0.$$

On the other hand, by Chebyshev's inequality and the strong type estimate for the Hardy-Littlewood maximal function, see Theorem 2.22, for $f \in L^p(\mathbb{R}^n)$, $1 < p < \infty$, we have

$$|\{x \in \mathbb{R}^n : M_\phi f(x) > \lambda\}| \leqslant \frac{1}{\lambda^p}\|Mf\|_p^p \leqslant \frac{1}{\lambda^p}\|f\|_p^p \quad \text{for every} \quad \lambda > 0. \tag{3.2}$$

Thus (3.2) holds for $1 \leqslant p < \infty$. The proof is based on these two estimates and is somewhat similar to the proof of Theorem 2.24.

Let $\eta > 0$. Since compactly supported continuous functions are dense in $L^p(\mathbb{R}^n)$, there exists $g \in C_0(\mathbb{R}^n)$ such that $\|f - g\|_p < \eta$. Since g is continuous, there exists

$\delta > 0$ such that $|g(x-y)-g(x)| < \eta$ whenever $|y| < \delta$. This implies

$$\begin{aligned}|(\phi_\varepsilon * g)(x) - ag(x)| &\leqslant \int_{\mathbb{R}^n} |g(x-y)-g(x)|\phi_\varepsilon(y)\,dy \\ &\leqslant \eta \int_{B(0,\delta)} \phi_\varepsilon(y)\,dy + 2\|g\|_\infty \int_{\mathbb{R}^n \setminus B(0,\delta)} \phi_\varepsilon(y)\,dy.\end{aligned}$$

By Lemma 3.16

$$\int_{B(0,\delta)} \phi_\varepsilon(y)\,dy \leqslant \|\phi_\varepsilon\|_1 = \|\phi\|_1$$

and

$$\int_{\mathbb{R}^n \setminus B(0,\delta)} \phi_\varepsilon(y)\,dy \to 0 \quad \text{as} \quad \varepsilon \to 0.$$

By letting $\eta \to 0$ and $\varepsilon \to 0$, we have

$$\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * g)(x) - ag(x)| = 0 \quad \text{for every} \quad x \in \mathbb{R}^n.$$

It follows that

$$\begin{aligned}&\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| \\ &\qquad \leqslant \limsup_{\varepsilon\to 0} |\phi_\varepsilon * (f-g)(x) - a(f-g)(x)| + \underbrace{\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * g)(x) - ag(x)|}_{=0} \\ &\qquad \leqslant M_\phi(f-g)(x) + a|(f-g)(x)|.\end{aligned}$$

Let

$$A_i = \left\{x \in \mathbb{R}^n : \limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| > \frac{1}{i}\right\}, \quad i = 1,2,\dots.$$

Then

$$A_i \subset \left\{x \in \mathbb{R}^n : M_\phi(f-g)(x) > \frac{1}{2i}\right\} \cup \left\{x \in \mathbb{R}^n : |f(x)-g(x)| > \frac{1}{2i}\right\}, \quad i = 1,2,\dots,$$

and by (3.2) and Chebyshev's inequality we have

$$\begin{aligned}|A_i| &\leqslant \left|\left\{x \in \mathbb{R}^n : M_\phi(f-g)(x) > \frac{1}{2i}\right\}\right| + \left|\left\{x \in \mathbb{R}^n : |f(x)-g(x)| > \frac{1}{2i}\right\}\right| \\ &\leqslant ci^p \|f-g\|_p^p + (2i)^p \|f-g\|_p^p \\ &= ci^p \|f-g\|_p^p \leqslant ci^p \eta^p, \quad i = 1,2,\dots,\end{aligned}$$

By letting $\eta \to 0$, we conclude $|A_i| = 0$ for every $i = 1,2,\dots$ and thus $|\bigcup_{i=1}^\infty A_i| \leqslant \sum_{i=1}^\infty |A_i| = 0$. This shows that

$$|\{x \in \mathbb{R}^n : \limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| > 0\}| = 0,$$

from which we conclude that

$$\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| = 0 \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

$\boxed{p=\infty}$ Let $f \in L^\infty(\mathbb{R}^n)$. We show that

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * f)(x) = af(x) \quad \text{for almost every} \quad x \in B(0,r),\, r>0.$$

Let $f_1 = f\chi_{B(0,r+1)}$ and $f_2 = f - f_1$. Then $f_1 \in L^1(\mathbb{R}^n)$ and by the beginning of the proof

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * f_1)(x) = af_1(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

We claim that

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * f_2)(x) = 0 \quad \text{for almost every} \quad x \in B(0,r),\, r>0.$$

To see this, let $x \in B(0,r)$ and $|y|<1$. Then $x-y \in B(0,r+1)$ and thus $f_2(x-y)=0$. This implies

$$\begin{aligned}
|(\phi_\varepsilon * f_2)(x)| &= \left|\int_{\mathbb{R}^n} f_2(x-y)\phi_\varepsilon(y)\,dy\right| \\
&= \left|\int_{\mathbb{R}^n\setminus B(0,1)} f_2(x-y)\phi_\varepsilon(y)\,dy\right| \\
&= \|f_2\|_\infty \int_{\mathbb{R}^n\setminus B(0,1)} \phi_\varepsilon(y)\,dy \to 0 \quad \text{as} \quad \varepsilon\to 0.
\end{aligned}$$

□

Remark 3.21. If $f \in L^\infty(\mathbb{R}^n)$ is continuous at x, then

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * f)(x) = af(x).$$

Moreover, if $f \in L^\infty(\mathbb{R}^n)$ is uniformly continuous, the convergence is uniform. (Exercise)

3.5 Convergence in L^p

The previous theorem asserts that the convolution approximations converge almost everywhere, but in general almost everywhere convergence does not imply convergence in L^p. However, the next result shows that this is true as well.

Theorem 3.22. Assume that $\phi \in L^1(\mathbb{R}^n)$, $a = \int_{\mathbb{R}^n} \phi(x)\,dx$ and $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p < \infty$. Then

$$\lim_{\varepsilon\to 0} \|\phi_\varepsilon * f - af\|_p = 0.$$

THE MORAL: The convolution approximation converges in L^p.

WARNING: The result does not hold true for $p=\infty$. In this case the corresponding claim claim is the following: If $f \in L^\infty(\mathbb{R}^n)$ is uniformly continuous, then $\phi_\varepsilon * f \to af$ uniformly in $\mathbb{R}^n$, that is,

$$\lim_{\varepsilon\to 0} \|\phi_\varepsilon * f - af\|_\infty = 0.$$

Remark 3.23. If ϕ is the function in Examples 3.15 (1), by the Lebesgue differentiation theorem

$$\lim_{\varepsilon\to 0}\phi_\varepsilon * f(x) = f(x) \quad \text{for almost every} \quad x\in\mathbb{R}^n.$$

The theorem above asserts that $\phi_\varepsilon * f \to f$ in $L^p(\mathbb{R}^n)$ as well.

Proof. By Lemma 3.16 (1), we have

$$af(x) = f(x)\int_{\mathbb{R}^n}\phi(y)\,dy = f(x)\int_{\mathbb{R}^n}\phi_\varepsilon(y)\,dy = \int_{\mathbb{R}^n} f(x)\phi_\varepsilon(y)\,dy.$$

Thus

$$\begin{aligned}|(f*\phi_\varepsilon)(x) - af(x)| &= \left|\int_{\mathbb{R}^n}(f(x-y)-f(x))\phi_\varepsilon(y)\,dy\right| \\ &\leqslant \int_{\mathbb{R}^n}|f(x-y)-f(x)||\phi_\varepsilon(y)|^{1/p}|\phi_\varepsilon(y)|^{1/p'}\,dy. \\ &\left(\frac{1}{p}+\frac{1}{p'}=1,\ 1<p<\infty, \text{the case } p=1 \text{ is an exercise}\right)\end{aligned}$$

By Fubini's theorem and Hölder's inequality

$$\begin{aligned}&\int_{\mathbb{R}^n}\left|(f*\phi_\varepsilon)(x)-af(x)\right|^p\,dx \\ &\leqslant \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p|\phi_\varepsilon(y)|\,dy\right)\left(\int_{\mathbb{R}^n}|\phi_\varepsilon(y)|\,dy\right)^{p/p'}\,dx \\ &= \|\phi\|_1^{p/p'}\int_{\mathbb{R}^n}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p\,dx\right)dy,\end{aligned}$$

where

$$\int_{\mathbb{R}^n}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p\,dx\right)dy = \int_{B(0,r)}\dots\,dy + \int_{\mathbb{R}^n\setminus B(0,r)}\dots\,dy$$

Let $\eta>0$. By Theorem 3.6 there exists $r>0$ such that

$$\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p\,dx < \frac{\eta}{2(\|\phi\|_1^p+1)} \quad \text{for every} \quad y\in B(0,r).$$

This shows that

$$\begin{aligned}&\|\phi\|_1^{p/p'}\int_{B(0,r)}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p\,dx\right)dy \\ &\qquad\leqslant \frac{\eta}{2}\frac{\|\phi\|_1^{p/p'+1}}{\|\phi\|_1^p+1}\leqslant\frac{\eta}{2}. \qquad \left(\frac{p}{p'}+1=p\right)\end{aligned}$$

By Lemma 3.16 (2), there exists $\varepsilon'>0$ such that

$$\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\,dy < \frac{\eta}{2^{p+2}(\|f\|_p^p\|\phi\|_1^{p/p'}+1)}, \quad \text{as} \quad 0<\varepsilon<\varepsilon'.$$

This shows that

$$\begin{aligned}
\|\phi\|_1^{p/p'} \int_{\mathbb{R}^n\setminus B(0,r)} |\phi_\varepsilon(y)| \left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|^p\,dx \right) dy \\
&\leqslant 2^{p+1}\|\phi\|_1^{p/p'}\|f\|_p^p \int_{\mathbb{R}^n\setminus B(0,r)} |\phi_\varepsilon(y)|\,dy \\
&\quad \left(|f(x-y)-f(x)|^p \leqslant 2^p\left(|f(x-y)|^p+|f(x)|^p\right)\right) \\
&< \frac{\eta}{2}, \quad \text{as} \quad 0<\varepsilon<\varepsilon'.
\end{aligned}$$

Thus

$$\|f*\phi_\varepsilon - af\|_p^p < \frac{\eta}{2}+\frac{\eta}{2}=\eta.$$

□

Remark 3.24. An examination of the proof above shows that a more general result holds as well. Let $\phi_i \in L^1(\mathbb{R}^n)$, $i=1,2,\dots$, is a sequence with the properties

(1) $\displaystyle\lim_{i\to\infty}\int_{\mathbb{R}^n}\phi_i(x)\,dx = a$,

(2) $\displaystyle\sup_i \int_{\mathbb{R}^n}|\phi_i(x)|\,dx<\infty$ and

(3) $\displaystyle\lim_{i\to\infty}\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_i(x)|\,dx = 0$ for every $r>0$.

Then

$$\lim_{\varepsilon\to 0}\|\phi_\varepsilon * f - af\|_p = 0.$$

Note that here ϕ_i do not have to be nonnegative or given by the formula for the approximate identity.

3.6 Smoothing properties

For a positive integer m, let $C^m(\mathbb{R}^n)$ denote the class of functions $f:\mathbb{R}^n\to\mathbb{R}$, whose partial derivatives

$$D^\alpha f = \frac{\partial^{\alpha_1+\dots+\alpha_n} f}{\partial x_1^{\alpha_1}\dots\partial x_n^{\alpha_n}}$$

up to in including those of order m exist and are continuous. The subset of $C^m(\mathbb{R}^n)$ with functions of compact support is denoted by $C_0^m(\mathbb{R}^n)$. Moreover, $C^\infty(\mathbb{R}^n)$ is the class of functions which have continuous partial derivatives of all orders, that is,

$$C^\infty(\mathbb{R}^n) = \bigcap_{m=1}^{\infty} C^m(\mathbb{R}^n),$$

and $C_0^\infty(\mathbb{R}^n)$ is the corresponding class of functions with a compact support. Note that there are such a functions.

Example 3.25. The function Let $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi(x) = \begin{cases} \exp\left(\frac{1}{|x|^2-1}\right), & x \in B(0,1), \\ 0, & x \in \mathbb{R}^n \setminus B(0,1), \end{cases}$$

in Example 3.15 (2) belongs to $C_0^\infty(\mathbb{R}^n)$ (exercise).

Hint: First define $h : \mathbb{R} \to \mathbb{R}$,

$$h(t) = \begin{cases} 0, & t \leqslant 0, \\ \exp(-1/t), & t > 0. \end{cases}$$

Then $h \in C^\infty(\mathbb{R})$. Prove by induction that $h^{(m)}(t) = P_m(1/t)\exp(-1/t)$ for $t > 0$, where P_m is a polynomial of degree $2m$. Then prove by induction that $h^{(m)}(0) = 0$.) Now $\phi(x) = h(1-|x|^2)$ belongs to $C^\infty(\mathbb{R}^n)$ as a composed function of two functions in $C^\infty(\mathbb{R}^n)$. Moreover, if $|x| \geqslant 1$, then $1-|x|^2 \leqslant 0$ and thus $h(1-|x|^2) = 0$. Therefore this function belongs to $C_0^\infty(\mathbb{R}^n)$.

Theorem 3.26. If $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p \leqslant \infty$ and $\phi \in C_0^\infty(\mathbb{R}^n)$, then $f * \phi_\varepsilon \in C^\infty(\mathbb{R}^n)$ and

$$D^\alpha(f * \phi_\varepsilon)(x) = (f * D^\alpha \phi_\varepsilon)(x)$$

for every $x \in \mathbb{R}^n$, $\varepsilon > 0$, $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$.

THE MORAL: Thus the convolution inherits the smoothness of the mollifier, since we differentiate under the integral sign. This is justified by the Lebesgue differentiation theorem.

WARNING: In general $f * \phi_\varepsilon \notin C_0^\infty(\mathbb{R}^n)$, that is, the convolution approximation does not have a compact support.

Remark 3.27. Theorem 3.22 implies that $C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, when $1 \leqslant p < \infty$.

Proof. We observe that $f * \phi_\varepsilon$ is a continuous function (exercise). Let $e_i = (0, \dots, 1, \dots, 0)$ be the standard ith basis vector in $\mathbb{R}^n$, $i = 1, \dots, n$, $0 < |h| < 1$, $h \in \mathbb{R}$. Then

$$\frac{(f * \phi_\varepsilon)(x + he_i) - (f * \phi_\varepsilon)(x)}{h} = \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} \frac{1}{h} \left[\phi\left(\frac{x + he_i - y}{\varepsilon} \right) - \phi\left(\frac{x - y}{\varepsilon} \right) \right] f(y)\, dy$$

CLAIM:

$$\frac{1}{h} \left[\phi\left(\frac{x + he_i - y}{\varepsilon} \right) - \phi\left(\frac{x - y}{\varepsilon} \right) \right] \quad \to \quad \frac{1}{\varepsilon} \frac{\partial \phi}{\partial x_i} \left(\frac{x - y}{\varepsilon} \right) \quad \text{as} \quad h \to 0.$$

Reason. Let

$$\varphi(x) = \phi\left(\frac{x - y}{\varepsilon} \right).$$

Then

$$\frac{\partial \varphi}{\partial x_i}(x) = \frac{1}{\varepsilon} \frac{\partial \phi}{\partial x_i} \left(\frac{x - y}{\varepsilon} \right).$$

■

Next we derive a bound so that we may apply the Lebesgue dominated conevergence theorem. By the fundamental theorem of calculus

$$\begin{aligned}\varphi(x+he_i)-\varphi(x) &= \int_0^h \frac{\partial\varphi}{\partial t}(x+te_i)\,dt \\ &= \int_0^h D\varphi(x+te_i)\cdot e_i\,dt \qquad \left(D\varphi = \left(\frac{\partial\varphi}{\partial x_1},\dots,\frac{\partial\varphi}{\partial x_n}\right) = \text{the gradient}\right) \\ &= \int_0^h \frac{\partial\varphi}{\partial x_i}(x+te_i)\,dt,\end{aligned}$$

from which we have

$$\begin{aligned}|\varphi(x+he_i)-\varphi(x)| &\leqslant \int_0^{|h|}\left|\frac{\partial\varphi}{\partial x_i}(x+te_i)\right|dt \\ &= \frac{1}{\varepsilon}\int_0^{|h|}\left|\frac{\partial\phi}{\partial x_i}\left(\frac{x+te_i-y}{\varepsilon}\right)\right|dt \leqslant \frac{|h|}{\varepsilon}\|D\phi\|_\infty\end{aligned}$$

Define

$$K = \left\{y\in\mathbb{R}^n : \frac{x-y}{\varepsilon}\in\operatorname{supp}\phi \text{ or } \frac{x+he_i-y}{\varepsilon}\in\operatorname{supp}\phi,\ 0<|h|<1\right\}.$$

Since $\operatorname{supp}\phi$ is compact, we see that K is a bounded set. By the Lebesgue dominated convergence theorem

$$\begin{aligned}\frac{\partial(f*\phi_\varepsilon)}{\partial x_i}(x) &= \lim_{h\to0}\frac{(f*\phi_\varepsilon)(x+he_i)-(f*\phi_\varepsilon)(x)}{h} \\ &= \lim_{h\to0}\frac{1}{\varepsilon^n}\int_K\frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right)-\phi\left(\frac{x-y}{\varepsilon}\right)\right]f(y)\,dy \\ &= \frac{1}{\varepsilon^n}\int_K\lim_{h\to0}\frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right)-\phi\left(\frac{x-y}{\varepsilon}\right)\right]f(y)\,dy \\ &\qquad\qquad \left(\text{LDCT, } \frac{1}{\varepsilon}\|D\phi\|_\infty|f|\in L^1(K)\right) \\ &= \frac{1}{\varepsilon^n}\int_K\frac{1}{\varepsilon}\frac{\partial\phi}{\partial x_i}\left(\frac{x-y}{\varepsilon}\right)f(y)\,dy \\ &= \int_K\frac{\partial\phi_\varepsilon}{\partial x_i}(x-y)f(y)\,dy = \left(\frac{\partial\phi_\varepsilon}{\partial x_i}*f\right)(x).\end{aligned}$$

Since this partial derivative is given by a similar convolution as in the definition of $f*\phi_\varepsilon$ itself, it is a continuous function. By induction it follows that $f*\phi_\varepsilon$ possesses continuous partial derivatives of all orders. ☐

Next we show that every function in $L^p(\mathbb{R}^n)$ can be approximated by compactly supported smooth functions, when $1\leqslant p<\infty$. This result does not hold true for $p=\infty$. This is simply because the uniform limit of continuous functions is itself continuous.

Remark 3.28. The closure of $C_0^\infty(\mathbb{R}^n)$ in $L^\infty(\mathbb{R}^n)$ is the subspace of $C(\mathbb{R}^n)$ consisting of functions satisfying

$$\lim_{|x|\to\infty} f(x) = 0.$$

(Exercise)

Theorem 3.29. $C_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$.

Proof. Assume $f \in L^p(\mathbb{R}^n)$ and let $\eta > 0$. Theorem 3.3 shows that $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ so that there exists $g \in C_0(\mathbb{R}^n)$ such that $\|f-g\|_{L^p(\mathbb{R}^n)} < \eta/2$. Let ϕ_ε be the standard mollifier in Example 3.15 (2). Theorem 3.26 shows that $g * \phi_\varepsilon \in C^\infty(\Omega)$. CLAIM: $\operatorname{supp}(g*\phi_\varepsilon)$ is compact.

Reason. If $(g*\phi_\varepsilon)(x) \neq 0$, then there exists $y \in \mathbb{R}^n$ such that $g(y)\phi_\varepsilon(x-y) \neq 0$, which implies that $g(y) \neq 0$ and $\phi_\varepsilon(x-y) \neq 0$. If $g(y) \neq 0$, then $y \in \operatorname{supp} g$ and we denote $K = \operatorname{supp} g$. If $\phi_\varepsilon(x-y) \neq 0$, then $|x-y| \leqslant \varepsilon$. Thus

$$K_\varepsilon = \{y \in \mathbb{R}^n : \operatorname{dist}(x,K) \leqslant \varepsilon\}$$

is a compact set and $(g*\phi_\varepsilon)(x) = 0$ for every $x \in \mathbb{R}^n \setminus K_\varepsilon$. This implies that $g*\phi_\varepsilon$ has a compact support. ■

By Theorem 3.22 there exists $\varepsilon' > 0$ such that

$$\|g-(g*\phi_\varepsilon)\|_p < \frac{\eta}{2} \quad \text{when} \quad 0 < \varepsilon < \varepsilon'.$$

Thus

$$\|f-(g*\phi_\varepsilon)\|_p \leqslant \|f-g\|_p + \|g-(g*\phi_\varepsilon)\|_p < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$ □

3.7 The Poisson kernel

We consider an example from the partial differential equations. Assume that $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p < \infty$. Let $P : \mathbb{R}^n \to \mathbb{R}$,

$$P(x) = c(n)\frac{1}{(1+|x|^2)^{(n+1)/2}}$$

be the Poisson kernel, where $c(n) = \Gamma((n+1)/2)/\pi^{(n+1)/2}$ is chosen such that

$$\int_{\mathbb{R}^n} P(x)\,dx = 1.$$

Then

$$P_\varepsilon(x) = \frac{1}{\varepsilon^n}P\left(\frac{x}{\varepsilon}\right) = c(n)\frac{\varepsilon}{(|x|^2+\varepsilon^2)^{(n+1)/2}}, \quad \varepsilon > 0,$$

is an approximation of the identity and we may apply the theory developed above. By Young's theorem, the Poisson integral of f

$$u(x,\varepsilon) = (f*P_\varepsilon)(x) = \int_{\mathbb{R}^n} P_\varepsilon(x-y)f(y)\,dy$$

is well defined and

$$\|f*P_\varepsilon\|_p \leqslant \|f\|_p\|P_\varepsilon\|_1 = \|f\|_p \quad \text{for every} \quad \varepsilon > 0.$$

Theorem 3.20 implies that

$$\lim_{\varepsilon\to 0}(f * P_\varepsilon)(x) = f(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

Theorem 3.26 shows that the function $x \mapsto u(x,\varepsilon) = (f * P_\varepsilon)(x)$ belongs to $C^\infty(\mathbb{R}^n)$ for every fixed $\varepsilon > 0$. Moroever, the function u is a solution to the Laplace equation in the upper half space

$$\mathbb{R}^{n+1}_+ = \{(x_1,\dots,x_n,\varepsilon) \in \mathbb{R}^{n+1} : \varepsilon > 0\},$$

that is,

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \dots + \frac{\partial^2 u}{\partial x_n^2} + \frac{\partial^2 u}{\partial \varepsilon^2} = 0 \quad \text{in} \quad \mathbb{R}^{n+1}_+.$$

Thus $u(x,\varepsilon) = (f * P_\varepsilon)(x)$ is a solution to the Dirichlet problem

$$\begin{cases} \Delta u = 0 & \text{in} \quad \mathbb{R}^{n+1}_+, \\ u = f & \text{on} \quad \partial\mathbb{R}^{n+1}_+ = \mathbb{R}^n, \end{cases}$$

in the sense that

$$\lim_{\varepsilon\to 0} u(x,\varepsilon) = f(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

Moreover, Theorem 3.22 shows that $u(x,\varepsilon) \to f(x)$ in $L^p(\mathbb{R}^n)$ as $\varepsilon \to 0$.

Note also, that by Theorem 3.19, there exists a constant c such that

$$\sup_{\varepsilon>0} |(f * P_\varepsilon)(x)| \leqslant cMf(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

THE MORAL: This gives a method to define and study a solution to a Dirichlet problem in the upper half space for boundary values that only belong to L^p. In particular, the boundary values do not have to be continuous or bounded. On the other hand, this gives another point of view to the convolution approximations. They can be seen as extensions of functions to the upper half space.

Remark 3.30. Let μ be a finite Radon measure on $\mathbb{R}^n$. The convolution of μ with a function $f \in L^1(\mathbb{R}^n;\mu)$ is defined as

$$(f * \mu)(x) = \int_{\mathbb{R}^n} f(x-y)\, d\mu(y).$$

It can be shown that

$$\|P_\varepsilon * \mu\|_1 \leqslant \mu(\mathbb{R}^n) \quad \text{and} \quad \lim_{\varepsilon\to 0} \|P_\varepsilon * \mu\|_1 = \mu(\mathbb{R}^n).$$

Moreover,

$$\lim_{\varepsilon\to 0} \int_{\mathbb{R}^n} (P_\varepsilon * \mu)(x) f(x)\, dx = \int_{\mathbb{R}^n} f(x)\, d\mu(x) \quad \text{for every} \quad f \in C_0(\mathbb{R}^n).$$

(Exercise). This means that the measures $(P_\varepsilon * \mu)(x)f(x)\,dx$ converge weakly to μ as $\varepsilon \to 0$. We shall discuss the weak convergence of measures later. Note that this holds, in particular, when μ is Dirac's delta. In this case we obtain the fundamental solution, which is the Poisson kernel itself.

Derivatives of measures are very useful tools in representing measures as integrals with respect to another measure. The Radon-Nikodym theorem is a version of the fundamental theorem of calculus for measures. It has applications not only in analysis but also in probability theory. Differentiation of measures also extend the Lebesgue differentiation theorem for more general Radon measures. A powerful Besicovitch covering theorem is used in the arguments.

4 Differentiation of measures

There exists a useful differentiation theory for measures which has similar features as the differentiation theory for real functions. The first problem is to find a way to define the derivative of measures and to show that it exists.

4.1 Covering theorems

Let us recall the definition of a Radon measure from the measure and integration theory.

Definition 4.1. Let μ be an outer measure on $\mathbb{R}^n$.

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

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

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

THE MORAL: The Lebesgue outer measure is a Radon measure. General 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. This will be discussed in this course.

For an arbitrary Radon measure μ on $\mathbb{R}^n$, there is no systematic way to control $\mu(B(x,5r))$ in terms of $\mu(B(x,r))$. The measure μ is called doubling, if there is a constant c such that

$$\mu(B(x,2r)) \leqslant c\mu(B(x,r)) \quad \text{for every} \quad x \in \mathbb{R}^n, r > 0.$$

If μ is doubling, then

$$\begin{aligned}\mu(B(x,5r)) &\leqslant c\mu(B(x,5r/2)) \leqslant c^2\mu(B(x,5r/4))\\ &\leqslant c^3\mu(B(x,5r/8)) \leqslant c^3\mu(B(x,r)) \quad \text{for every} \quad x\in\mathbb{R}^n, r>0.\end{aligned}$$

Let A is a bounded subset of $\mathbb{R}^n$ and suppose that for every $x\in A$ there is a ball $B(x,r_x)$ with the radius $r_x>0$ possibly depending on the point x. By the Vitali covering theorem, see Theorem 2.16, we have a countable subcollection of pairwise disjoint balls $B(x_i,r_i)$, $i=1,2,\dots$, dilates of which covers the union of the original balls. Thus

$$\begin{aligned}\mu(A) &\leqslant \mu\Big(\bigcup_{x\in A} B(x,r_x)\Big) \leqslant \mu\Big(\bigcup_{i=1}^{\infty} B(x_i,5r_i)\Big) \leqslant \sum_{i=1}^{\infty}\mu(B(x_i,5r_i))\\ &= c^3\sum_{i=1}^{\infty}\mu(B(x_i,r_i)) = c^3\mu\Big(\bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) \leqslant c^3\mu\Big(\bigcup_{x\in A} B(x,r_x)\Big).\end{aligned}$$

This shows that for a doubling measure we can use similar covering arguments as for the Lebesgue measure.

However, for a general Radon measure Theorem 2.16 is not useful. We need a covering theorem that does not require us to enlarge the balls, but allows the balls to have overlap. The claim is purely geometric and it will be an important tool to prove other covering theorems.

Theorem 4.2 (Besicovitch covering theorem). There exist integers P and Q, depending only on n, with the following properties. Let $A\subset\mathbb{R}^n$ be a bounded set and let $\mathscr{F}$ be a collection of closed balls $B(x,r)$ such that every point of A is a center of some ball in $\mathscr{F}$.

(1) There is a countable subcollection of balls $B(x_i,r_i)\in\mathscr{F}$, $i=1,2,\dots$, such that they cover the set A, that is,

$$A\subset\bigcup_{i=1}^{\infty}B(x_i,r_i)$$

and that every point of $\mathbb{R}^n$ belongs to at most P balls $B(x_i,r_i)$, that is,

$$\chi_A\leqslant\sum_{i=1}^{\infty}\chi_{B(x_i,r_i)}\leqslant P.$$

(2) There are subcollections $\mathscr{F}_1,\dots,\mathscr{F}_Q\subset\mathscr{F}$ such that each $\mathscr{F}_k$ consists of countably many pairwise disjoint balls in $\mathscr{F}$ and

$$A\subset\bigcup_{k=1}^{Q}\bigcup_{B(x_i,r_i)\in\mathscr{F}_k}B(x_i,r_i).$$

THE MORAL: Property (1) asserts that the subcollection covers the set of center points of the original balls and that the balls in the subcollection have bounded overlap. Property (2) asserts that the subcollection can be distributed in a finite number of subcollections of disjoint balls.

Let A is a bounded subset of $\mathbb{R}^n$ and suppose that for every $x \in A$ there is a ball $B(x, r_x)$ with the radius $r_x > 0$ possibly depending on the point x. By the Besicovitch covering theorem, we have a countable subcollection of balls $B(x_i, r_i)$, $i = 1, 2, \dots$, which covers the union of the original balls. Thus

$$\begin{aligned}\mu(A) &\leqslant \mu\Big(\bigcup_{i=1}^{\infty} B(x_i, r_i)\Big) \leqslant \sum_{i=1}^{\infty} \mu(B(x_i, r_i)) \\ &\leqslant P\mu\Big(\bigcup_{i=1}^{\infty} B(x_i, r_i)\Big) \leqslant P\mu\Big(\bigcup_{x\in A} B(x, r_x)\Big).\end{aligned}$$

Example 4.3. Let μ be the Radon measure on $\mathbb{R}^2$ defined by

$$\mu(A) = |\{x \in \mathbb{R} : (x, 0) \in A\}|,$$

where $|\cdot|$ denotes the one-dimensional Lebesgue measure. The collection

$$\mathscr{F} = \{B((x, y), y) : x \in \mathbb{R}, 0 < y < \infty\}$$

of closed balls covers the set $A = \{(x, 0) : x \in \mathbb{R}\}$, but for any countable subcollection we have

$$\mu\Big(A \cap \bigcup_{i=1}^{\infty} B_i\Big) = 0.$$

THE MORAL: The previous example shows that it is essential in the Besicovitch covering theorem, that every point of A is (more or less) a center of some ball in the collection. In particular, it is not enough, that every point of A belongs to a ball in the collection.

We need a couple of lemmas in the proof of the Besicovitch covering theorem.

Lemma 4.4. If $x, y \in \mathbb{R}^n$, $0 < |x| < |x - y|$ and $0 < |y| < |x - y|$, then

$$\left| \frac{x}{|x|} - \frac{y}{|y|} \right| \geqslant 1.$$

THE MORAL: This means that the angle between points x and y is at least $60°$.

Reason. Since

$$\begin{aligned}\left| \frac{x}{|x|} - \frac{y}{|y|} \right|^2 &= \left\langle \frac{x}{|x|} - \frac{y}{|y|}, \frac{x}{|x|} - \frac{y}{|y|} \right\rangle \\ &= \frac{\langle x, x\rangle}{|x|^2} - \frac{\langle x, y\rangle}{|x||y|} - \frac{\langle y, x\rangle}{|x||y|} + \frac{\langle y, y\rangle}{|y|^2} \\ &= 2 - 2\frac{\langle x, y\rangle}{|x||y|},\end{aligned}$$

we have

$$\begin{aligned}\left|\frac{x}{|x|}-\frac{y}{|y|}\right|\geqslant 1 \iff \quad & 2-2\frac{\langle x,y\rangle}{|x||y|}\geqslant 1\\ \iff \quad & \frac{\langle x,y\rangle}{|x||y|}\leqslant\frac{1}{2}\\ \iff \quad & \cos\angle(x,y)\leqslant\frac{1}{2}\\ \iff \quad & \angle(x,y)\geqslant 60^\circ\end{aligned}$$

■

Proof. We may assume that $n=2$, since there is a plane containing x, y and the origin. Moreover, we may assume that $x=(x_1,0,\dots,0)$. If $x\notin B(y,|y|)$ and $y\notin B(x,|x|)$, then by plane geometry $y_1\leqslant\frac{x_1}{2}$. Since

$$\cos\alpha=\frac{\frac{|x|}{2}}{|x|}=\frac{1}{2}$$

we conclude that $\angle(x,y)\geqslant 60^\circ$ (Figure required). Another way to prove the lemma is to use the cosine theorem. □

The following lemma is the core of the proof of the Besicovitch covering theorem.

Lemma 4.5. There exists a constant $N=N(n)$ such that if $x_1,\dots,x_k\in\mathbb{R}^n$ and $r_1,\dots,r_k>0$ such that

(1) $x_j\notin B(x_i,r_i)$, whenever $i\neq j$ and
(2) $\bigcap_{i=1}^k B(x_i,r_i)\neq\emptyset$,

then $k\leqslant N(n)$.

T H E M O R A L : Condition (1) asserts that the center of every ball belongs only to that ball of which center it is and (2) asserts that all the balls intersect at some point. The claim is that there can be only finitely many such balls.

Remark 4.6. For example infinite dimensional Hilbert space l^2 does not have the property above, so that it is some kind finite dimensionality condition.

Reason. Let e_i, $i=1,2,\dots$, be the standard orthonormal basis of l^2, that is, the ith term of e_i is one and all other terms are zero. Then every closed ball $B(e_i,1)$, $i=1,2,\dots$, contains the origin and every e_i belongs to only that ball of which center it is.

This example also shows that the Besicovitch covering theorem does not hold in l^2. Indeed, if we remove any ball $B(e_i,1)$, then the center e_i is not covered by the other balls. Moreover, the balls do not have bounded overlap at the origin. ■

Proof. We may assume that $0 \in \bigcap_{i=1}^{k} B(x_i, r_i)$. Then (1) implies that $x_i \neq 0$ for every $i = 1, \dots, k$. To see this, assume on the contrary that $x_{i_0} = 0$ for some $i_0 = 1, \dots, k$. Then $x_{i_0} \in B(x_i, r_i)$ for every $i = 1, 2, \dots, k$, which is not possible. We have

$$0 < |x_i| < r_i < |x_i - x_j|, \quad j \neq i.$$

Lemma 4.4 implies

$$\left| \frac{x_i}{|x_i|} - \frac{x_j}{|x_j|} \right| \geqslant 1, \quad j \neq i \tag{4.1}$$

SInce $\partial B(0,1) \subset \mathbb{R}^n$ is a compact set, it can be covered by finitely many balls $B(y_i, \frac{1}{2})$ with $y_i \in \partial B(0,1)$, $i = 1, \dots, N(n)$.

Then $k \leqslant N$, since otherwise for some indices $i, j \leqslant k$, $i \neq j$, the points $\frac{x_i}{|x_i|}$ and $\frac{x_j}{|x_j|}$ would belong to the same ball $B(y_{i_0}, \frac{1}{2})$ with $i_0 \leqslant N$. This implies

$$\left| \frac{x_i}{|x_i|} - \frac{x_j}{|x_j|} \right| < 1,$$

which contradicts (4.1). □

Now we are ready for the proof of the Besicovitch covering theorem. This proof is technical and can be omitted in the first reading.

Proof. (1) Step 1 Since A is bounded and or every $x \in A$ there exists $B(x, r_x) \in \mathscr{B}$, we may assume that

$$M_1 = \sup\{r_x : x \in A\} < \infty.$$

(If $r_x > 2diamA$, then the single ball $B(x, r_x)$ satisfies the required properties.) Choose $x_1 \in A$ such that $r_{x_1} \geqslant \frac{M_1}{2}$. Then we choose recursively

$$x_{j+1} \in A \setminus \bigcup_{i=1}^{j} B(x_i, r_{x_j}) \quad \text{such that} \quad r_{x_{j+1}} \geqslant \frac{M_1}{2},$$

as long as this is possible. Since $|x_i - x_j| \geqslant \frac{M_1}{2}$, $i \neq j$, and $\frac{M_1}{2} \leqslant r_{x_i} \leqslant M_1$, we conclude that the balls $B\left(x_i, \frac{M_1}{4}\right)$, $i = 1, 2, \dots$, are disjoint. Then $B\left(x_i, \frac{M_1}{4}\right) \subset B(x, R)$ with $R = \operatorname{diam}(A) + M$ and $x \in A$. This implies

$$\sum_{i=1}^{k} \left| B\left(x_i, \frac{M_1}{4}\right) \right| = \left| \bigcup_{i=1}^{k} B\left(x_i, \frac{M_1}{4}\right) \right| \leqslant |B(x, R)|.$$

On the other hand,

$$\sum_{i=1}^{k} \left| B\left(x_i, \frac{M_1}{4}\right) \right| = k|B(0,1)| \left(\frac{M_1}{4} \right)^n$$

which implies

$$k \leqslant \left(\frac{4}{M_1} \right)^n \frac{|B(x,R)|}{|B(0,1)|} < \infty$$

and consequently there are only finitely many points x_i, $i = 1, \dots, k_1$.

Denote

$$M_2 = \sup\left\{ r_x : x \in A \setminus \bigcup_{i=1}^{k_1} B(x, r_{x_i}) \right\} < \infty.$$

Choose

$$x_{k_1+1} \in A \setminus \bigcup_{i=1}^{k_1} B(x, r_{x_i}) \quad \text{such that} \quad r_{x_{k_1+1}} \geqslant \frac{M_2}{2}$$

and recursevily

$$x_{j+1} \in A \setminus \bigcup_{i=1}^{j} B(x_i, r_{x_i}) \quad \text{such that} \quad r_{x_{j+1}} \geqslant \frac{M_2}{2}.$$

By the construction, we have $M_2 \leqslant \frac{M_1}{2}$. Again we obtain finitely many points as above. By continuing this way, we obtain a countably, or finitely, many

(1) indices $0 = k_0 < k_1 < k_2 < \dots$,

(2) numbers M_i such that $M_{i+1} \leqslant \dfrac{M_i}{2}$,

(3) balls $B(x_i, r_{x_i}) \in \mathscr{B}$ and

(4) classes of indices $I_j = \{k_{j-1}+1, \dots, k_j\}$, $j = 1, 2, \dots$.

We shall show that the collection $B(x_i, r_{x_i})$, $i = 1, 2, \dots$, has the desired properties.

CLAIM:

$$\frac{M_j}{2} \leqslant r_{x_i} \leqslant M_j \leqslant \frac{M_{j-1}}{2}, \quad \text{when} \quad i \in I_j, \tag{4.2}$$

$$x_{j+1} \in A \setminus \bigcup_{i=1}^{j} B(x_i, r_{x_i}) \quad \text{and} \tag{4.3}$$

$$x_i \in A \setminus \bigcup_{m \neq k} \bigcup_{j \in I_m} B(x_j, r_{x_j}), \quad \text{when} \quad i \in I_k. \tag{4.4}$$

Reason. The first two properties (4.2) and (4.3) follow from the construction. To prove (4.4), assume that $i \in I_k$, $m \neq k$ and $j \in I_m$. If $m < k$, then by (4.3) we have $x_i \notin B(x_j, r_{x_j})$. If $k < m$, then (4.2) and $m - 1 \geqslant k$ imply

$$r_{x_j} \leqslant M_m \leqslant \frac{M_{m-1}}{2} \leqslant \frac{M_k}{2} \leqslant r_{x_i}.$$

Thus (4.3) implies $x_j \notin B(x_i, r_{x_i})$ and consequently $x_i \notin B(x_j, r_{x_j})$. ■

Observe that

$$\begin{aligned} (4.3) \implies & M_i \leqslant 2^{1-i} M_1, \; i = 1, 2, \dots \\ \implies & M_i \to 0, \; i \to \infty \\ \implies & r_{x_i} \to 0, \; i \to \infty. \end{aligned}$$

CLAIM: $A \subset \bigcup_{i=1}^{\infty} B(x_i, r_{x_i})$.

Reason. Assume, on the contrary, that there exists $x \in A \setminus \bigcup_{i=1}^{\infty} B(x_i, r_{x_i})$. Then there exists j such that $\frac{M_j}{2} \leqslant r_x \leqslant M_j$, which implies that $x \in \bigcup_{i \in I_j} B(x_i, r_{x_i})$. This is a contradiction. ∎

Step 2 We shall show that every $x \in \mathbb{R}^n$ belongs to at most $P(n) = 16^n N(n)$ balls, where $N(n)$ is as in Lemma 4.5 Assume that $x \in \bigcap_{i=1}^{p} B(x_{m_i}, r_{x_{m_i}})$.

If $B_1, \dots, B_s$ are balls in the collection $\{B(x_{m_i}, r_{x_{m_i}})\}_{i=1,\dots,p}$ with the property that each ball belongs to a different class of indices I_j. Property (4.4) implies that $x \in \bigcap_{k=1}^{s} B_k$ and every ball B_k does not contain the center of any other ball B_j with $k \neq j$. Lemma 4.5 implies $s \leqslant N(n)$ and

$$\sharp \left\{ j : I_j \cap \{m_i : i = 1, \dots, p\} \neq \emptyset \right\} \leqslant N(n).$$

In other words, the indices m_i can belong to at most $N(n)$ classes of indices I_j.

CLAIM: $\sharp\{I_j \cap \{m_i : i = 1, \dots, p\}\} \leqslant 16^n$, $j = 1, 2, \dots$.

Reason. Fix j and denote

$$I_j \cap \{m_i : i = 1, \dots, p\} = \{l_1, \dots l_q\}$$

Properties (4.2) and (4.3) imply $B(x_{l_i}, \frac{1}{4} r_{x_{l_i}})$, $i = 1, \dots q$, are pairwise disjoint and they are contained in the ball $B(x, 2M_j)$. Thus

$$q |B(0,1)| \left(\frac{M_j}{8} \right)^n \leqslant \sum_{i=1}^{q} \left| B\left(x_{l_i}, \frac{M_j}{4} \right) \right| \leqslant |B(x, 2M_j)| = |B(0,1)| (2M_j)^n$$

This implies $q = 16^n$. ∎

(2) Let $B(x_i, r_{x_i})$, $i = 1, 2, \dots$, be the collection of balls in the claim (1) of the Besicovitch covering. Since $M_j \to 0$, $j \to \infty$, for every $\varepsilon > 0$ there are only a finite number of balls $B(x_i, r_{x_i})$ such that $r_{x_i} \geqslant \varepsilon$. Thus we may assume that $r_{x_1} \geqslant r_{x_2} \geqslant \dots$. Denote $B_i = B(x_i, r_{x_i})$, $i = 1, 2, \dots$. Let $B_{1,1} = B_1$ and inductively $B_{1,j+1} = B_k$, where k is the smallest index, for which

$$B_k \cap \bigcup_{i=1}^{j} B_{1,i} = \emptyset.$$

Continuing this way, we obtain a countable (or finite) subcollection

$$\mathscr{B}_1 = \{B_{1,1}, B_{1,2}, \dots\},$$

which consists of pairwise disjoint balls. If $\bigcup_{i=1}^{\infty} B_{1,i}$ does not cover the set A, we choose $B_{2,1} = B_k$, where k is the smallest index for which $B_k \notin \mathscr{B}_1$.

Inductively, let $B_{2,j+1} = B_k$, where k is the smallest index for which

$$B_k \cap \bigcup_{i=1}^{j} B_{2,j} = \emptyset.$$

This gives subcollections $\mathscr{B}_1, \mathscr{B}_2, \dots$ consisting of pairwise disjoint balls.

CLAIM: $A \subset \bigcup_{k=1}^{m} \bigcup_{i=1}^{\infty} B_{k,i}$ with $m = 4^n P + 1$.

Reason. We show that, if there exists $x \in A \setminus \bigcup_{k=1}^m \bigcup_{i=1}^\infty B_{k,i}$, then $m \leqslant 4^n P$. Since $A \subset \bigcup_{i=1}^\infty B_i$, there exists i such that $x \in B_i = B(x_i, r_{x_i})$. Then $B_i \notin \mathscr{B}_k$, $k \leqslant m$ and, by the definition of $\mathscr{B}_k$, there exists B_{k,i_k} such that $B_{k,i_k} \cap B_i \neq \emptyset$ and $r_{x_i} \leqslant r_{x_{i_k}}$ for every $k \leqslant m$. Thus for every $k \leqslant m$ there exists a ball

$$B'_k \subset B(x_i, 2r_{x_i}) \cap B_{k,i_k}$$

such that the radius of B'_k is $r_i/2$. By (1), each point in $\mathbb{R}^n$ belongs to at most P balls B_{k,i_k}, $k = 1, \dots, m$. This holds for subballs B'_k as well. This implies

$$\sum_{k=1}^m \chi_{B'_k} \leqslant P \chi_{\bigcup_{k=1}^m B'_k}$$

and consequently

$$\begin{aligned} 2^n r_i^n |B(0,1)| = |B(x_i, 2r_{x_i})| &\geqslant \left| \bigcup_{k=1}^m B'_k \right| \qquad (B'_k \subset B(x_i, 2r_i)) \\ &= \int_{\mathbb{R}^n} \chi_{\bigcup_{k=1}^m B'_k} \, dx \geqslant \frac{1}{P} \int_{\mathbb{R}^n} \sum_{k=1}^m \chi_{B'_k} \, dx \\ &= \frac{1}{P} \sum_{k=1}^m |B'_k| = \frac{m}{P} |B(0,1)| \left(\frac{r_{x_i}}{2} \right)^n. \end{aligned}$$

This shows that $m \leqslant 4^n P$. ■

Remarks 4.7:

(1) The assumption that A is bounded can be replaced with

$$\sup\{r : B(x,r) \in \mathscr{F}\} < \infty$$

in the Besicovitch covering theorem 4.2.

(2) The balls in the Besicovitch covering theorem 4.2 need not be closed.

(3) The balls can be replaced, for example, by cubes in the Besicovitch covering theorem 4.2.

We take another look at the Vitali covering theorem.

Theorem 4.8 (Infinitesimal Vitali covering theorem). Let μ be a Radon measure on $\mathbb{R}^n$, $A \subset \mathbb{R}^n$ and $\mathscr{F}$ a collection of closed balls such that each point of A is a center of arbitrarily small balls $\mathscr{F}$, that is,

$$\inf\{r > 0 : B(x,r) \in \mathscr{F}\} = 0 \quad \text{for every} \quad x \in A.$$

Then there are disjoint balls $B(x_i, r_i) \in \mathscr{F}$, $i = 1, 2, \dots$, such that

$$\mu\Big(A \setminus \bigcup_{i=1}^\infty B(x_i, r_{x_i})\Big) = 0.$$

THE MORAL: The infinitesimal Vitali covering theorem implies that every open set can be exhausted by a countable many disjoint balls up to a set of measure zero.

Proof. We may assume that $\mu(A) > 0$, because otherwise the claim is clear. Assume first that A is bounded. By approximation properties of measurable sets for a Radon measure, there exists an open set $G \supset A$ such that

$$\mu(G) \leqslant \left(1 + \frac{1}{4Q}\right)\mu(A).$$

By the Besicovitch covering theorem, there are subcollections $\mathscr{F}_1, \dots, \mathscr{F}_Q$ such that the balls in each $\mathscr{F}_k$ are disjoint and

$$A \subset \bigcup_{k=1}^{Q} \bigcup_{B(x_i,r_i)\in\mathscr{F}_k} B(x_i,r_i) \subset G.$$

This implies

$$\mu(A) \leqslant \sum_{k=1}^{Q} \mu\Big(\bigcup_{\mathscr{F}_k} B(x_i,r_i)\Big).$$

Thus there exists k such that

$$\mu(A) \leqslant Q\mu\Big(\bigcup_{\mathscr{F}_k} B(x_i,r_i)\Big).$$

Reason. If $\mu(A) > Q\mu\left(\bigcup_{\mathscr{F}_k} B(x_i,r_i)\right)$ for every $k = 1, \dots, Q$, then

$$Q\mu(A) = \sum_{k=1}^{Q} \mu(A) > Q \sum_{k=1}^{Q} \mu\Big(\bigcup_{\mathscr{F}_k} B(x_i,r_i)\Big).$$

This is impossible. ■

There exists a finite subcollection $\mathscr{F}_1' \subset \mathscr{F}_k$ such that

$$\mu(A) \leqslant 2Q\mu\Big(\bigcup_{\mathscr{F}_1'} B(x_i,r_i)\Big).$$

Let

$$A_1 = A \setminus \bigcup_{\mathscr{F}_1'} B(x_i,r_i).$$

Then

$$\begin{aligned}
\mu(A_1) &\leqslant \mu\Big(G \setminus \bigcup_{\mathscr{F}_1'} B(x_i,r_i)\Big) \\
&= \mu(G) - \mu\Big(\bigcup_{\mathscr{F}_1'} B(x_i,r_i)\Big) \leqslant \left(1 + \frac{1}{4Q} - \frac{1}{2Q}\right)\mu(A) \\
&= \left(1 - \frac{1}{4Q}\right)\mu(A) = \gamma\mu(A), \quad \gamma = 1 - \frac{1}{4Q} < 1.
\end{aligned}$$

In practice, this means that balls in $\mathscr{F}_1'$ cover a certain percentage of A in the sense of measure.

Then we apply the same argument to the collection

$$\mathscr{F}' = \left\{ B(x,r) \in \mathscr{F} : B(x,r) \cap \Big(\bigcup_{\mathscr{F}_1'} B(x_i,r_i) \Big) = \emptyset \right\}.$$

Note that A_1 is a subset of the open set $G \setminus \bigcup_{\mathscr{F}_1'} B(x_i,r_i)$. Thus there exists an open set G_1 such that

$$A_1 \subset G_1 \subset G \setminus \bigcup_{\mathscr{F}_1'} B(x_i,r_i) \quad \text{and} \quad \mu(G_1) \leqslant \Big(1 + \frac{1}{4Q}\Big)\mu(A_1).$$

Thus there exists a finite subcollection $\mathscr{F}_2'$ such that

$$\Big(\bigcup_{\mathscr{F}_1'} B(x_i,r_i) \Big) \cap \Big(\bigcup_{\mathscr{F}_2'} B(x,r) \Big) = \emptyset$$

and

$$\mu(A_2) \leqslant \gamma \mu(A_1), \quad \text{where} \quad A_2 = A \setminus \bigcup_{\mathscr{F}_1',\mathscr{F}_2'} B(x,r).$$

By continuing this process, we obtain

$$\mu\Big(A \setminus \bigcup_{\mathscr{F}_1' \cup \ldots \cup \mathscr{F}_k'} B(x,r) \Big) \leqslant \gamma^k \mu(A)$$

and the result follows by letting $k \to \infty$, since $\gamma < 1$ and $\mu(A) < \infty$.

In order to remove the assumption that A bounded, we use the fact that $\mu(\partial B(0,r)) > 0$ for at most countably many radii $r > 0$, if μ is a Radon measure. Hence we may choose the radii $0 < r_1 < r_2 < \ldots$ such that $r_k \to \infty$ as $k \to \infty$ and $\mu(\partial B(0,r_k)) = 0$ for every $k = 1,2,\ldots$

Denote $A_1 = \{x \in \mathbb{R}^n : |x| < r_1\}$,

$$A_k = \{x \in \mathbb{R}^n : r_{k-1} < |x| < r_k\}, \quad k = 2,3,\ldots$$

and

$$\mathscr{F}^k = \{B(x,r) \in \mathscr{F} : B(x,r) \subset A_k,\ x \in A\}.$$

The claim follows by applying the proof above for the sets A_k and the coverings $\mathscr{F}^k$, $k = 1,2,\ldots$ □

4.2 The Lebesgue differentiation theorem for Radon measures

It is not immediately clear how to define the derivative of measures. Let $f : [a,b] \to [0,\infty]$ be a nonnegative integrable function and define $F(x) = \int_{[a,x]} f(y)\,dy$. Then by Theorem 2.33, $F'(x) = f(x)$ for almost every $x \in [a,b]$. Let us rewrite this in another way. Let $\mu(A) = \int_A f(y)\,dy$ for every Lebesgue measurable set $A \subset \mathbb{R}$ and denote the one-dimensional Lebesgue measure by ν. Then

$$\frac{F(x+r)-F(x)}{r} = \frac{1}{r}\int_{[x,x+r]} f(y)\,dy = \frac{\mu([x,x+r])}{\nu([x,x+r])}.$$

Thus

$$F'(x) = \lim_{r\to 0}\frac{\mu([x,x+r])}{\nu([x,x+r])} = f(x) \quad \text{for almost every} \quad x \in [a,b].$$

This suggest the following definition for the derivative of measures.

Definition 4.9. Let μ and ν be Radon measures on $\mathbb{R}^n$. The upper derivative of ν with respect to μ is

$$\overline{D}_\mu \nu(x) = \limsup_{r\to 0}\frac{\nu(B(x,r))}{\mu(B(x,r))}$$

and the lower derivative of ν with respect to μ is

$$\underline{D}_\mu \nu(x) = \liminf_{r\to 0}\frac{\nu(B(x,r))}{\mu(B(x,r))}.$$

We use the convention that $\overline{D}_\mu \nu(x) = \infty$ and $\underline{D}_\mu \nu(x) = \infty$, if $\mu(B(x,r)) = 0$ for some $r > 0$. At the points where the limit exists, we define the derivative of ν with respect to μ as

$$D_\mu \nu(x) = \overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x) < \infty.$$

Examples 4.10:

(1) Let $A \subset \mathbb{R}^n$ be μ-measurable. Then by the measure theory, $\nu = \mu\lfloor A$ is a Radon measure and

$$\lim_{r\to 0}\frac{\nu(B(x,r))}{\mu(B(x,r))} = \lim_{r\to 0}\frac{\mu(A\cap B(x,r))}{\mu(B(x,r))}$$

measures the density of A at x.

(2) Assume $f \in L^1(\mathbb{R}^n;\mu)$ and define $\nu(A) = \int_A |f|\,d\mu$ for every μ-measurable set $A \subset \mathbb{R}^n$. Then

$$\lim_{r\to 0}\frac{\nu(B(x,r))}{\mu(B(x,r))} = \lim_{r\to 0}\frac{1}{\mu(B(x,r))}\int_{B(x,r)} |f|\,d\mu$$

is the limit of the integral averages as in the Lebesgue differentiation theorem.

Lemma 4.11. $\overline{D}_\mu \nu$, $\underline{D}_\mu \nu$ and $D_\mu \nu$ are Borel measurable.

Proof. CLAIM: $\limsup_{y\to x} \mu(B(y,r)) \leqslant \mu(B(x,r))$ for every $x \in \mathbb{R}^n$.

Reason. By an approximation result for measurable sets, there exists $G \supset B(x,r)$ such that $\mu(G) < \mu(B(x,r)) + \varepsilon$. It follows that $B(y,r) \subset G$, if $|x-y| < \operatorname{dist}(B(x,r),\mathbb{R}^n \setminus G)/2$. Thus

$$\mu(B(y,r)) \leqslant \mu(G) < \mu(B(x,r)) + \varepsilon, \quad \text{if} \quad |x-y| < \frac{1}{2}\operatorname{dist}(B(x,r),\mathbb{R}^n \setminus G).$$

This implies

$$\limsup_{y\to x}\mu(B(y,r))\leqslant\mu(B(x,r))$$

and thus $x\mapsto\mu(B(x,r))$ is upper semicontinuous. Similarly $x\mapsto\nu(B(x,r))$ us upper semicontinuous and consequently the functions are Borel measurable. ■

CLAIM:

$$\overline{D}_\mu\nu(x)=\limsup_{\substack{r\to0\\ r\in\mathbb{Q}_+}}\frac{\nu(B(x,r))}{\mu(B(x,r))}.$$

Reason. Since $B(x,r)$ is a closed ball,

$$B(x,r)=\bigcap_{i=1}^{\infty}B\left(x,r+\frac{1}{i}\right)\quad\text{and}\quad\mu(B(x,r+1))<\infty,$$

we have

$$\mu(B(x,r))=\mu\left(\bigcap_{i=1}^{\infty}B\left(x,r+\frac{1}{i}\right)\right)=\lim_{i\to\infty}\mu\left(B\left(x,r+\frac{1}{i}\right)\right).$$

This implies that μ and ν are continuous from right and that we may replace the limes superior with a limes superior over the rationals. Consequently, $\overline{D}_\mu\nu$ is a countable limes superior of Borel functions and hence it is a Borel function. The measurability of $\underline{D}_\mu\nu$ and $D_\mu\nu$ are proved in a similar manner (exercise). ■

The following result will be an extremely useful tool in our analysis.

Theorem 4.12. Assume that μ and ν are Radon measures in $\mathbb{R}^n$, $A\subset\mathbb{R}^n$ and $0<t<\infty$.

(1) If $\underline{D}_\mu\nu(x)\leqslant t$ for every $x\in A$, then $\nu(A)\leqslant t\mu(A)$.
(2) If $\overline{D}_\mu\nu(x)\geqslant t$ for every $x\in A$, then $\nu(A)\geqslant t\mu(A)$.

THE MORAL: These inequalities give distribution set estimates

$$\nu(\{x\in\mathbb{R}^n:\underline{D}_\mu\nu(x)\leqslant t\})\leqslant t\mu(\mathbb{R}^n)$$

and

$$\mu(\{x\in\mathbb{R}^n:\overline{D}_\mu\nu(x)\geqslant t\})\leqslant\frac{1}{t}\nu(\mathbb{R}^n).$$

which are Chebyshev type inequalities for Radon measures.

Remark 4.13. The set $A\subset\mathbb{R}^n$ does not necessarily have to be measurable, compare to the Vitali covering theorem (Theorem 4.8).

Proof. (1) If $\mu(A)=\infty$, the claim is clear, so that we may assume $\mu(A)<\infty$. Let $\varepsilon>0$. Threre exists an open $G\supset A$ such that $\mu(G)<\mu(A)+\varepsilon$. Since $\underline{D}_\mu\nu(x)\leqslant t$ for every $x\in A$ there is an arbitrarily small $r>0$ such that

$$\nu(B(x,r))\leqslant(t+\varepsilon)\mu(B(x,r))\quad\text{and}\quad B(x,r)\subset G.$$

By the Vitali covering theorem (Theorem 4.8), there are pairwise disjoint balls $B(x_i,r_i)$, $i=1,2,\dots$, such that

$$\nu\Big(A\setminus\bigcup_{i=1}^{\infty}B(x_i,r_i)\Big)=0.$$

Thus

$$\begin{aligned}\nu(A)&\leqslant \nu\Big(A\cap\bigcup_{i=1}^{\infty}B(x_i,r_i)\Big)+\underbrace{\nu\Big(A\setminus\bigcup_{i=1}^{\infty}B(x_i,r_i)\Big)}_{=0}\\ &\leqslant \sum_{i=1}^{\infty}\nu(B(x_i,r_i))\leqslant (t+\varepsilon)\sum_{i=1}^{\infty}\mu(B(x_i,r_i)\\ &\leqslant (t+\varepsilon)\mu\Big(\bigcup_{i=1}^{\infty}B(x_i,r_i)\Big)\quad\text{(the balls are disjoint)}\\ &\leqslant (t+\varepsilon)\mu(G)\leqslant (t+\varepsilon)(\mu(A)+\varepsilon)\quad\text{for every}\quad \varepsilon>0.\end{aligned}$$

Letting $\varepsilon\to 0$, we have $\nu(A)\leqslant t\mu(A)$.

(2) (Exercise) □

Theorem 4.14. If μ and ν are Radon measures on $\mathbb{R}^n$, then the derivative $D_\mu\nu(x)$ exists and is finite for μ-almost every $x\in\mathbb{R}^n$.

THE MORAL: This is a version of the Lebesgue differentiation theorem for general Radon measures.

Proof. Step 1 $\overline{D}_\mu\nu=\underline{D}_\mu\nu$ μ-almost everywhere in $\mathbb{R}^n$ or equivalently

$$\mu(\{x\in\mathbb{R}^n:\overline{D}_\mu\nu(x)>\underline{D}_\mu\nu(x)\})=0.$$

Reason. Let $i,k\in\{1,2,\dots\}$, $p,q\in\mathbb{Q}$ with $p<q$. Define

$$A_{k,p,q}=\{x\in B(0,k):\underline{D}_\mu\nu(x)\leqslant p<q\leqslant\overline{D}_\mu\nu(x)\}$$

and

$$A_{k,i}=\{x\in B(0,k):\overline{D}_\mu\nu(x)\geqslant i\}.$$

Theorem 4.12 implies

$$q\mu(A_{k,p,q})\leqslant\nu(A_{k,p,q})\leqslant p\mu(A_{k,p,q}),$$

from which it follows that

$$\mu(A_{k,p,q})\leqslant\frac{p}{q}\mu(A_{k,p,q})\leqslant\underbrace{\mu(B(0,k))}_{<\infty}.$$

Since $p<q$, we conclude that $\mu(A_{k,p,q})=0$. Thus

$$\begin{aligned}\mu(\{x\in B(0,k):&\overline{D}_\mu\nu(x)>\underline{D}_\mu\nu(x)\})\\ &=\mu\Big(\bigcup_{\substack{0<p<q\\ p,q,\in\mathbb{Q}}}A_{k,p,q}\Big)\\ &\leqslant\sum_{\substack{0<p<q\\ p,q\in\mathbb{Q}}}\mu(A_{k,p,q})=0\end{aligned}$$

and consequently

$$\mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) > \underline{D}_\mu \nu(x)\}) \leqslant \sum_{k=1}^{\infty} \mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) > \underline{D}_\mu \nu(x)\}) = 0.$$

Since $\overline{D}_\mu \nu \geqslant \underline{D}_\mu \nu$ always, we conclude that $\overline{D}_\mu \nu = \underline{D}_\mu \nu$ μ-almost everywhere in $\mathbb{R}^n$. ∎

Step 2 $\overline{D}_\mu \nu < \infty$ μ-almost everywhere in $\mathbb{R}^n$ or equivalently

$$\mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) = \infty\}) = 0.$$

Reason. Theorem 4.12 implies

$$\mu(A_{k,i}) \leqslant \frac{1}{i}\nu(A_{k,i}) \leqslant \frac{1}{i}\underbrace{\nu(B(0,k))}_{<\infty}.$$

Thus

$$\mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}) \leqslant \mu(A_{k,i}) \leqslant \frac{1}{i}\nu(B(0,k)) \quad \text{for every} \quad i,k = 1,2,\dots.$$

By letting $i \to \infty$, we have

$$\mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}) = 0 \quad \text{for every} \quad k = 1,2,\dots.$$

Therefore

$$\begin{aligned}\mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) = \infty\}) &= \mu\Big(\bigcup_{k=1}^{\infty} \{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}\Big) \\ &\leqslant \sum_{k=1}^{\infty} \mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}) = 0.\end{aligned}$$ ∎

4.3 The Radon-Nikodym theorem

Assume that μ and ν are Radon measures on $\mathbb{R}^n$. Let f be a nonnegative μ-measurable function and define $\nu(A) = \int_A f\,d\mu$, where A μ-measurable. Then ν is a measure with the property that $\mu(A) = 0$ implies $\nu(A) = 0$. Conversely, if ν is a Radon measure on $\mathbb{R}^n$, does there exist a μ-measurable function f such that $\nu(A) = \int_A f\,d\mu$ for every μ-measurable set A? The Radon-Nikodym theorem shows that this is the case if ν is absolutely continuous with respect to μ.

Definition 4.15. A outer measure ν is absolutely continuous with respect to another outer measure μ, if $\mu(A) = 0$ implies $\nu(A) = 0$. In this case we write $\nu \ll \mu$.

THE MORAL: $\nu \ll \mu$ means that ν is small if μ is small. When we are dealing with more than one measure, the term almost everywhere becomes ambiguous and we have to specify almost everywhere with respect μ or ν. If $\nu \ll \mu$ and a property holds μ-almost everywhere, then it also holds ν-almost everywhere.

Examples 4.16:

(1) Let μ be the Lebesgue measure and ν be the Dirac measure at the origin,

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

Then $\mu(\{0\}) = 0$, but $\nu(\{0\}) = 1$. Thus ν is not absolutely continuous with respect to μ.

(2) Let $f \in L^1(\mathbb{R}^n;\mu)$ and define $\nu(A) = \int_A |f|\,d\mu$, where A μ-measurable. Then $\mu(A) = 0$ implies $\nu(A) = \int_A |f|\,d\mu = 0$. Thus $\nu \ll \mu$.

Remarks 4.17:

(1) It is often useful, in particular in connection with integrals, to use the following ε, δ version of absolute continuity: If ν is a finite measure, then $\nu \ll \mu$ if and only if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\nu(A) < \varepsilon$ for every μ-measurable set A with $\mu(A) < \delta$. In particular, if $f \in L^1(\mathbb{R}^n;\mu)$, then for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\int_A |f|\,d\mu < \varepsilon$ for every μ-measurable set A with $\mu(A) < \delta$. (Exercise)

(2) It is easy to verify that that the relation $\ll$ is reflexive ($\mu \ll \mu$) and transitive ($\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_3$ imply $\mu_1 \ll \mu_3$.)

The following theorem on absolutely continuous measures is very important. It shows that differentiation of measures and integration are inverse operations. It has applications in the identification of continuous linear functionals on L^p, $1 \leqslant p < \infty$. Moreover, a general version of the theorem is applied in the construction of the conditional expectation in the probability theory.

Theorem 4.18 (Radon-Nikodym theorem). Assume that μ and ν are Radon measures on $\mathbb{R}^n$ and $\nu \ll \mu$. Then

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

for every μ-measurable set $A \subset \mathbb{R}^n$.

THE MORAL: The Radon-Nikodym theorem asserts that ν can be expressed as an integral with respect to μ and the Radon-Nikodym derivative $D_\mu \nu$ can be computed by differentiating ν with respect to μ. This is a version of the fundamental theorem of calculus for Radon measures.

Remark 4.19. Note that $D_\mu \nu$ does not have to be integrable. In fact, $D_\mu \nu \in L^1(\mathbb{R}^n;\mu)$ if and only if $\nu(\mathbb{R}^n) < \infty$.

Proof. Step 1 Since A is a μ-measurable set and μ is a Borel regular measure, there exists a Borel set $B \supset A$ such that $\mu(B \setminus A) = 0$. Since $\nu \ll \mu$, we have $\nu(B \setminus A) = 0$ and thus $B \setminus A$ is ν-measurable. This implies that $A = B \setminus (B \setminus A)$ is ν-measurable.

Step 2 Define

$$Z = \{x \in A : D_\mu \nu(x) = 0\},$$

$$I = \{x \in A : \overline{D}_\mu \nu(x) = \infty\}$$

and

$$N = \{x \in A : \overline{D}_\mu \nu(x) \neq \underline{D}_\mu \nu(x)\}.$$

By Theorem 4.14 $\mu(I) = 0$ and $\mu(N) = 0$. Since $\nu \ll \mu$, we also have $\nu(I) = 0$ and $\nu(N) = 0$. Theorem 4.12 implies

$$\nu(Z \cap B(0,i)) \leqslant t \underbrace{\mu(Z \cap B(0,i))}_{<\infty} \quad \text{for every} \quad i = 1,2,\dots, \quad 0 < t < \infty.$$

By letting $t \to 0$, we obtain $\nu(Z \cap B(0,i)) = 0$ for every $i = 1,2,\dots$. Thus

$$\nu(Z) = \nu\Big(\bigcup_{i=1}^{\infty}(Z \cap B(0,i))\Big) \leqslant \sum_{i=1}^{\infty} \underbrace{\nu(Z \cap B(0,i))}_{=0} = 0.$$

Define

$$A_i = \left\{x \in A : t^i \leqslant D_\mu \nu(x) < t^{i+1}\right\}, \quad i \in \mathbb{Z}, \quad 1 < t < \infty,$$

Then $A \setminus \bigcup_{i=-\infty}^{\infty} A_i \subset Z \cup I \cup N$ and

$$\nu\Big(A \setminus \bigcup_{i=-\infty}^{\infty} A_i\Big) \leqslant \nu(Z \cup I \cup N) \leqslant \nu(Z) + \nu(I) + \nu(N) = 0.$$

Step 3 :

$$\begin{aligned}
\nu(A) &= \nu\Big(A \cap \bigcup_{i=-\infty}^{\infty} A_i\Big) + \underbrace{\nu\Big(A \setminus \bigcup_{i=-\infty}^{\infty} A_i\Big)}_{=0} \leqslant \sum_{i=-\infty}^{\infty} \nu(A_i) \\
&\leqslant \sum_{i=-\infty}^{\infty} t^{i+1}\mu(A_i) \qquad \text{(Theorem 4.12 (1))} \\
&= t \sum_{i=-\infty}^{\infty} t^i \mu(A_i) \\
&\leqslant t \sum_{i=-\infty}^{\infty} \int_{A_i} D_\mu \nu \, d\mu \qquad (D_\mu \nu \geqslant t^i \text{ in } A_i) \\
&= t \int_A D_\mu \nu \, d\mu.
\end{aligned}$$

In the last equality we used the facts that $A = \bigcup_{i=-\infty}^{\infty} A_i \cup (Z \cup I \cup N)$, where the sets are disjoint and

$$\int_A D_\mu \nu \, d\mu = \int_{\bigcup_{i=-\infty}^{\infty} A_i} D_\mu \nu \, d\mu + \underbrace{\int_Z D_\mu \nu \, d\mu}_{=0} + \underbrace{\int_I D_\mu \nu \, d\mu}_{=0, (\mu(I)=0)} + \underbrace{\int_N D_\mu \nu \, d\mu}_{=0, (\mu(N)=0)}.$$

On the other hand,

$$\begin{aligned}
\nu(A) &\geqslant \nu\Big(\bigcup_{i=-\infty}^{\infty} A_i\Big) \\
&= \sum_{i=-\infty}^{\infty} \nu(A_i) \quad (A_i \text{ disjoint}) \\
&\geqslant \sum_{i=-\infty}^{\infty} t^i \mu(A_i) \qquad (\text{Theorem 4.12 (2)}) \\
&= \frac{1}{t} \sum_{i=-\infty}^{\infty} t^{i+1} \mu(A_i) \\
&\geqslant \frac{1}{t} \sum_{i=-\infty}^{\infty} \int_{A_i} D_\mu \nu \, d\mu \qquad (D_\mu \nu \leqslant t^{i+1} \text{ in } A_i) \\
&= \frac{1}{t} \int_A D_\mu \nu \, d\mu. \qquad (\text{as above})
\end{aligned}$$

Thus

$$\frac{1}{t} \int_A D_\mu \nu \, d\mu \leqslant \nu(A) \leqslant t \int_A D_\mu \nu \, d\mu, \quad 1 < t < \infty$$

and by letting $t \to 1$, we arrive at

$$\nu(A) = \int_A D_\mu \nu \, d\mu.$$

□

Remark 4.20. The Radon-Nikodym derivative has many properties reminiscent of standard derivatives. Let ν, μ and ζ be Radon measures on $\mathbb{R}^n$.

(1) if $\nu \ll \mu$ and f is a nonnegative μ-measurable function, then

$$\int_A f \, d\nu = \int_A f D_\mu \nu \, d\mu$$

for every measurable set A. (Exercise)
Hint: if g is a nonnegative μ-measurable function and $\nu(A) = \int_A g \, d\mu$ for every μ-measurable set $A \subset \mathbb{R}^n$, then for every nonnegative measurable function f we have

$$\int_A f \, d\nu = \int_A f g \, d\mu.$$

(2) If $\nu \ll \mu$ and $\zeta \ll \mu$, then $D_\mu(\nu + \zeta) = D_\mu \nu + D_\mu \zeta$ μ-almost everywhere.

(3) If $\nu \ll \zeta \ll \mu$, then $D_\mu \nu = D_\zeta \nu D_\mu \zeta$ μ-almost everywhere.

(4) If $\nu \ll \mu$ and $\mu \ll \nu$, then $D_\mu \nu = 1/D_\nu \mu$ μ-almost everywhere.

Remark 4.21. The Radon-Nikodym derivative is unique: If $f \in L^1_{\text{loc}}(\mathbb{R}^n;\mu)$ is a nonnegative function and define $\nu(A) = \int_A f \, d\mu$ for every μ-measurable set $A \subset \mathbb{R}^n$. Then $f = D_\mu \nu$ μ-almost everywhere. (Exercise)

Remark 4.22. The Radon-Nikodym theorem holds in a more general context: If μ is a σ-finite measure on X and ν is a σ-finite signed measure on X such that $\nu \ll \mu$.

Then there exists a real-valued measurable function f such that $\nu(A) = \int_A f\,d\mu$ for every measurable set $A \subset X$ with $|\nu|(A) < \infty$. If g is another function such that $\nu(A) = \int_A g\,d\mu$ for every measurable set $A \subset X$ with $|\nu|(A) < \infty$, then $f = g$ μ-almost everywhere. The function f above is called the Radon-Nikodym derivative of ν with respect to μ. However, in the general case there is no formula for the Radon-Nikodym derivative.

4.4 The Lebesgue decomposition

In this section we consider measures which are not necessarily absolutely continuous. The following definition describes an extreme form of non absolute continuity.

Definition 4.23. The Radon measures μ and ν are mutually singular, if there exists a Borel set $B \subset \mathbb{R}^n$ such that

$$\mu(\mathbb{R}^n \setminus B) = \nu(B) = 0.$$

In this case we write $\mu \perp \nu$.

THE MORAL: Mutually singular measures live on complementary sets.

Remark 4.24. Absolutely continuous and singular measures have the following general properties (exercise):

(1) If $\nu_1 \perp \mu$ and $\nu_2 \perp \mu$, then $(\nu_1 + \nu_2) \perp \mu$.

(2) If $\nu_1 \ll \mu$ and $\nu_2 \ll \mu$, then $(\nu_1 + \nu_2) \ll \mu$.

(3) If $\nu_1 \ll \mu$ and $\nu_2 \perp \mu$, then $\nu_1 \perp \nu_2$.

(4) If $\nu \ll \mu$ and $\nu \perp \mu$, then $\nu = 0$.

Theorem 4.25 (Lebesgue decomposition). Let μ and ν be Radon measures on $\mathbb{R}^n$.

(1) Then $\nu = \nu_a + \nu_s$, where ν_a and ν_s are Radon measures with $\nu_a \ll \mu$ and $\nu_s \perp \mu$.

(2) Furthermore, $D_\mu \nu = D_\mu \nu_a$ and $D_\mu \nu_s = 0$ μ-almost everywhere in $\mathbb{R}^n$ and

$$\nu(A) = \int_A D_\mu \nu_a \, d\mu + \nu_s(A)$$

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

TERMINOLOGY: We call ν_a the absolutely continuous part and ν_s the singular part of ν with respect to μ.

THE MORAL: A Radon measure can be split into an absolutely continuous and singular parts. The absolutely continuous part lives in the set where $\overline{D}_\mu \nu < \infty$ and the singular part in the set where $\overline{D}_\mu \nu = \infty$.

Proof. Step 1 $\mathbb{R}^n = \bigcup_{i=1}^{\infty} B(0,i)$,so that by considering the restrictions $\mu\lfloor B(0,i)$ and $\nu\lfloor B(0,i)$, $i = 1,2,\ldots$, we may assume $\mu(\mathbb{R}^n) < \infty$ and $\nu(\mathbb{R}^n) < \infty$.

Step 2 Define

$$\mathscr{F} = \{A \subset \mathbb{R}^n : A \text{ Borel}, \mu(\mathbb{R}^n \setminus A) = 0\}.$$

Then for every $i = 1,2,\ldots$ there exits $B_i \in \mathscr{F}$ such that

$$\nu(B_i) \leqslant \inf_{A\in\mathscr{F}} \nu(A) + \frac{1}{i}.$$

Note that the infimum is finite, since $\mathbb{R}^n \in \mathscr{F}$ and $\nu(\mathbb{R}^n) < \infty$. $B = \bigcap_{i=1}^{\infty} B_i$ is a Borel set and

$$\mu(\mathbb{R}^n \setminus B) = \mu\Big(\bigcup_{i=1}^{\infty}(\mathbb{R}^n \setminus B_i)\Big) \leqslant \sum_{i=1}^{\infty} \underbrace{\mu(\mathbb{R}^n \setminus B_i)}_{=0,\,B_i\in\mathscr{F}} = 0.$$

Thus $B \in \mathscr{F}$ and

$$\inf_{A\in\mathscr{F}} \nu(A) \leqslant \nu(B) \leqslant \inf_{A\in\mathscr{F}} \nu(A) + \frac{1}{i} \quad \text{for every} \quad i = 1,2,\ldots.$$

By letting $i \to \infty$ we conclude

$$\nu(B) = \inf_{A\in\mathscr{F}} \nu(A).$$

Thus B is the smallest possible set in the sense of ν measure such that the complement is of μ-measure zero. Define

$$\nu_a = \nu\lfloor B \quad \text{and} \quad \nu_s = \nu\lfloor(\mathbb{R}^n \setminus B).$$

By the properties of the restrictions of measures, ν_a and ν_s are Radon measures.

Step 3 CLAIM: $\nu_a \ll \mu$.

Reason. Suppose, on the contrary, that $A \subset B$, A is a Borel set, $\mu(A) = 0$, but $\nu(A) > 0$. Then

$$\mu(\mathbb{R}^n \setminus (B \setminus A)) = \mu(A \cup (\mathbb{R}^n \setminus B)) \leqslant \underbrace{\mu(A)}_{=0} + \underbrace{\mu(\mathbb{R}^n \setminus B)}_{=0,\,B\in\mathscr{F}} = 0$$

and thus $B \setminus A \in \mathscr{F}$. This implies

$$\nu(B \setminus A) = \nu(B) - \nu(A) < \nu(B) = \inf_{A\in\mathscr{F}} \nu(A). \qquad (0 < \nu(A) < \infty)$$

This shows that $\nu(A) = 0$, from which it follows that $\nu_a(A) = \nu(A \cap B) = 0$ and consequently $\nu_a \ll \mu$. ■

On the other hand,

$$\mu(\mathbb{R}^n \setminus B) = 0 = \nu(\underbrace{(\mathbb{R}^n \setminus B) \cap B}_{=\emptyset}) = \nu_s(B),$$

so that $\nu_s \perp \mu$.

Step 4 Let

$$C_i = \left\{ x \in \mathbb{R}^n : D_\mu \nu_s(x) \geqslant \frac{1}{i} \right\}, \quad i = 1, 2, \ldots.$$

Then

$$\begin{aligned}
\frac{1}{i}\mu(C_i) &= \frac{1}{i}(\mu(C_i \cap B) + \underbrace{\mu(C_i \cap (\mathbb{R}^n \setminus B))}_{=0,\, \mu(\mathbb{R}^n \setminus B)=0} \\
&\leqslant \nu_s(C_i \cap B) \qquad \text{(Theorem 4.12)} \\
&= \nu(\underbrace{(C_i \cap B) \cap (\mathbb{R}^n \setminus B)}_{=\emptyset}) = 0.
\end{aligned}$$

This shows that $\mu(C_i) = 0$ for every $i = 1, 2, \ldots$, and consequently

$$\mu(\{x \in \mathbb{R}^n : D_\mu \nu_s(x) > 0\}) = \mu\Big(\bigcup_{i=1}^{\infty} C_i\Big) \leqslant \sum_{i=1}^{\infty} \mu(C_i) = 0.$$

Thus $D_\mu \nu_s(x) = 0$ for μ-almost every $x \in \mathbb{R}^n$ and

$$\begin{aligned}
D_\mu \nu_a(x) &= \lim_{r \to 0} \frac{\nu_a(B(x,r))}{\mu(B(x,r))} \\
&= \lim_{r \to 0} \frac{\nu((B(x,r)) - \nu_s(B(x,r))}{\mu(B(x,r))} \\
&= D_\mu \nu(x) - D_\mu \nu_s(x) = D_\mu \nu(x) \quad \text{for } \mu\text{-almost every} \quad x \in \mathbb{R}^n.
\end{aligned}$$

The Radon-Nikodym theorem implies

$$\nu_a(A) = \int_A D_\mu \nu_a \, d\mu, \quad A \subset \mathbb{R}^n \, \mu\text{-measurable}.$$

Thus

$$\begin{aligned}
\nu(A) &= \nu_a(A) + \nu_s(A) \\
&= \int_A D_\mu \nu_a \, d\mu + \nu_s(A) \\
&= \int_A D_\mu \nu \, d\mu + \nu_s(A).
\end{aligned}$$

□

Remark 4.26. It can be shown that $\mu \perp \nu$ if and only if $D_\mu \nu = 0$ μ-almost everywhere. (Exercise)

Remark 4.27. The Lebesgue decomposition holds in a more general context: Let μ and ν be σ-finite signed measures on X. Then $\nu = \nu_a + \nu_s$, where ν_a and ν_s are σ-finite signed measures with $\nu_a \ll \mu$ and $\nu_s \perp \mu$. Moreover, this decomposition is unique.

4.5 Lebesgue and density points revisited

We shall prove a version of the Lebesgue differentiation theorem for an arbitrary Radon measure on $\mathbb{R}^n$.

Theorem 4.28 (Lebesgue differentiation theorem). Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\text{loc}}(\mathbb{R}^n;\mu)$. Then

$$\lim_{r\to 0} \frac{1}{\mu\big(B(x,r)\big)} \int_{B(x,r)} f\,d\mu = f(x)$$

for μ-almost every $x \in \mathbb{R}^n$.

Proof. Define

$$\nu^{\pm}(B) = \int_B f^{\pm}\,d\mu,$$

when $B \subset \mathbb{R}^n$ is a Borel set and

$$\nu^{\pm}(A) = \inf\{\nu^{\pm}(B) : A \subset B,\ B \text{ Borel}\}$$

for an arbitrary $A \subset \mathbb{R}^n$. Then ν^+ and ν^- are Radon measures and $\nu^{\pm} \ll \mu$ (exercise). The Radon-Nikodym theorem (Theorem 4.18) implies

$$\nu^{\pm}(A) = \int_A D_\mu \nu^{\pm}\,d\mu = \int_A f^{\pm}\,d\mu$$

for every μ-measurable set $A \subset \mathbb{R}^n$. This implies that $D_\mu \nu^{\pm} = f^{\pm}$ μ-almost everywhere in $\mathbb{R}^n$. Consequently,

$$\begin{aligned}
\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f\,d\mu &= \lim_{r\to 0} \frac{1}{\mu(B(x,r))} \left[\int_{B(x,r)} f^+\,d\mu - \int_{B(x,r)} f^-\,d\mu\right] \\
&= \lim_{r\to 0} \frac{1}{\mu(B(x,r))} \left[\nu^+(B(x,r)) - \nu^-(B(x,r))\right] \\
&= D_\mu \nu^+(x) - D_\mu \nu^-(x) \\
&= f^+(x) - f^-(x) = f(x)
\end{aligned}$$

for μ-almost every $x \in \mathbb{R}^n$. □

Corollary 4.29. Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\text{loc}}(\mathbb{R}^n;\mu)$. Then

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - f(x)|\,d\mu = 0$$

for μ-almost every $x \in \mathbb{R}^n$.

Proof. Let $\bigcup_{i=1}^{\infty}\{q_i\} = \mathbb{Q}$ be an enumeration of the rationals. By Theorem 4.28, for every $i = 1, 2, \ldots$ there exists $A_i \subset \mathbb{R}^n$ such that $\mu(A_i) = 0$ and

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - q_i|\,d\mu = |f(x) - q_i| \quad \text{for every} \quad x \in \mathbb{R}^n \setminus A_i.$$

Let $A=\bigcup_{i=1}^{\infty}A_i$. Then $\mu(A)\leqslant\sum_{i=1}^{\infty}\mu(A_i)=0$ and

$$\lim_{r\to 0}\frac{1}{\mu(B(x,r))}\int_{B(x,r)}|f-q_i|\,d\mu=|f(x)-q_i|\quad\text{for every}\quad x\in\mathbb{R}^n\setminus A.$$

Fix $x\in\mathbb{R}^n\setminus A$ and $\varepsilon>0$. Then there exists q_i such that $|f(x)-q_i|<\varepsilon/2$. This implies

$$\begin{aligned}\limsup_{r\to 0}&\frac{1}{\mu(B(x,r))}\int_{B(x,r)}|f-f(x)|\,d\mu\\&\leqslant\limsup_{r\to 0}\frac{1}{\mu(B(x,r))}\left[\int_{B(x,r)}|f-q_i|\,d\mu+\int_{B(x,r)}|q_i-f(x)|\,d\mu\right]\\&=|f(x)-q_i|+|f(x)-q_i|<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.\end{aligned}$$

□

Remarks 4.30:

(1) We have already seen in Example 2.28 that

$$\lim_{r\to 0}\frac{1}{\mu\big(B(x,r)\big)}\int_{B(x,r)}f\,d\mu=f(x)$$

does not necessarily imply

$$\lim_{r\to 0}\frac{1}{\mu(B(x,r))}\int_{B(x,r)}|f-f(x)|\,d\mu=0$$

at a given point $x\in\mathbb{R}^n$. The point in the proof above is that the previous equality holds for every function $f\in L^1_{\text{loc}}(\mathbb{R}^n;\mu)$ for μ-almost every $x\in\mathbb{R}^n$ and this implies the latter equality for μ-almost every $x\in\mathbb{R}^n$.

(2) In contrast with the proof of the Lebesgue differentiation theorem (Theorem 2.24) based on the Hardy-Littlewood maximal function, this proof does not depend on the density of compactly supported continuous functions in $L^1(\mathbb{R}^n;\mu)$.

We discuss a special case of the Lebesgue differentiability theorem. Let $A\subset\mathbb{R}^n$ a μ-measurable set and consider $f=\chi_A$. By the Lebesgue differentation theorem

$$\lim_{r\to 0}\frac{1}{\mu(B(x,r))}\int_{B(x,r)}\chi_A\,d\mu=\lim_{r\to 0}\frac{\mu(A\cap B(x,r))}{\mu(B(x,r))}=\chi_A(x)$$

for μ-almost every $x\in\mathbb{R}^n$. In particular,

$$\lim_{r\to 0}\frac{\mu(A\cap B(x,r))}{\mu(B(x,r))}=1\quad\text{for }\mu\text{-almost every}\quad x\in A$$

and

$$\lim_{r\to 0}\frac{\mu(A\cap B(x,r))}{\mu(B(x,r))}=0\quad\text{for }\mu\text{-almost every}\quad x\in\mathbb{R}^n\setminus A.$$

Thus the theory of density points extends to general Radon measures on $\mathbb{R}^n$.

In this chapter we show that Radon measures arise naturally in connection with linear functionals on compactly supported continuous functions. Moreover, we consider weak convergence of Radon measures and L^p functions and obtain useful compactness theorems.

5 Existence, convergence and compactness for Radon measures

Radon measures on $\mathbb{R}^n$ interact nicely with the Euclidean topology. Indeed, measurable sets can be approximated open sets from outside and compact sets from inside and integrable functions can be approximated by compactly supported continuous functions. In this chapter we show that certain linear functionals on compactly supported continuous functions are characterized by integrals with respect to Radon measures. This fact constitutes an important link between measure theory and functional analysis and it also provides a useful tool for constructing such measures.

5.1 The Riesz representation theorem for L^p spaces

A mapping $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ is a linear functional, if

$$L(af + bg) = aL(f) + bL(g)$$

for every $f, g \in L^p(\mathbb{R}^n)$ and $a, b \in \mathbb{R}$. The functional L is bounded, if there exists a constant $M < \infty$ such that

$$|L(f)| \leqslant M\|f\|_p \quad \text{for every} \quad f \in L^p(\mathbb{R}^n).$$

The norm of L is the smallest constant M for which the bound above holds, that is,

$$\begin{aligned}\|L\| &= \sup_{f\in L^p(\mathbb{R}^n),\|f\|_p\neq 0} \frac{|L(f)|}{\|f\|_p}\\ &= \sup_{f\in L^p(\mathbb{R}^n),\|f\|_p\neq 0} \frac{L(f)}{\|f\|_p}\\ &= \sup_{f\in L^p(\mathbb{R}^n),\|f\|_p\leqslant 1} |L(f)|.\end{aligned}$$

Recall, that the linear functional

$$L : L^p(\mathbb{R}^n)\to\mathbb{R} \quad \text{is continuous} \quad \Longleftrightarrow \quad L \text{ is bounded} \quad \Longleftrightarrow \quad \|L\|<\infty.$$

The space of bounded linear functionals on $L^p(\mathbb{R}^n)$ is called the dual space of $L^p(\mathbb{R}^n)$. The dual space is denoted by $L^p(\mathbb{R}^n)^*$. The main result of this section provides us with a representation for continuous linear functionals on $L^p(\mathbb{R}^n)$ with $1\leqslant p<\infty$. This is called the Riesz representation theorem and it gives a characterization for $L^p(\mathbb{R}^n)^*$ with $1\leqslant p<\infty$. We begin with the easier direction.

Theorem 5.1. Let $1\leqslant p\leqslant\infty$ and assume that μ is a Radon measure. Then for every $g\in L^{p'}(\mathbb{R}^n)$, the functional $L : L^p(\mathbb{R}^n)\to\mathbb{R}$,

$$L(f)=\int_{\mathbb{R}^n} fg\,d\mu$$

is linear and bounded and thus belongs to $L^{p*}(\mathbb{R}^n)$. Moreover, $\|L\|=\|g\|_{p'}$.

T H E   M O R A L : This shows that for every function $g\in L^{p'}(\mathbb{R}^n)$ there is a bounded linear functional $L : L^p(\mathbb{R}^n)\to\mathbb{R}$ with $\|L\|=\|g\|_{p'}$. With this interpretation $L^{p'}(\mathbb{R}^n)\subset L^p(\mathbb{R}^n)^*$.

Proof. The linearity follows from the linearity of the integral. If $\|g\|_{p'}=0$, the claim is clear. Hence we may assume that $\|g\|_{p'}>0$.

$\boxed{1<p<\infty}$ By Hölder's inequality

$$\begin{aligned}|L(f)| &= \left|\int_{\mathbb{R}^n} fg\,d\mu\right| \leqslant \int_{\mathbb{R}^n}|f||g|\,d\mu\\ &\leqslant \left(\int_{\mathbb{R}^n}|f|^p\,d\mu\right)^{1/p}\left(\int_{\mathbb{R}^n}|g|^{p'}\,d\mu\right)^{1/p'} = \|f\|_p\|g\|_{p'}.\end{aligned}$$

This implies

$$\|L\| = \sup_{f\in L^p(\mathbb{R}^n),\|f\|_p\leqslant 1}|L(f)| \leqslant \|g\|_{p'}<\infty.$$

On the other hand, the function $f=|g|^{p'/p}\operatorname{sign} g$ belongs $L^p(\mathbb{R}^n)$, since

$$\|f\|_p = \left(\int_{\mathbb{R}^n}|f|^p\,d\mu\right)^{1/p} = \left(\int_{\mathbb{R}^n}|g|^{p'}\,d\mu\right)^{1/p} = \|g\|_{p'}^{p'/p}<\infty.$$

Thus

$$|L(f)| = \int_{\mathbb{R}^n} |g|^{p'/p} \underbrace{g \operatorname{sign} g}_{=|g|} \, d\mu = \int_{\mathbb{R}^n} |g|^{p'} \, d\mu$$

$$= \left(\int_{\mathbb{R}^n} |g|^{p'} \, d\mu \right)^{1/p'} \left(\int_{\mathbb{R}^n} \underbrace{|g|^{p'}}_{=|f|^p} \, d\mu \right)^{1/p} = \|f\|_p \|g\|_{p'}$$

and

$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \neq 0} \frac{|L(f)|}{\|f\|_p} \geqslant \|g\|_{p'}.$$

Since the inequality in the reverse direction holds always, we have $\|L\| = \|g\|_{p'}$.

$\boxed{p = \infty}$ Assume that $g \in L^1(\mathbb{R}^n)$. Then

$$|L(f)| \leqslant \int_{\mathbb{R}^n} |f||g| \, d\mu \leqslant \|f\|_\infty \|g\|_1,$$

which implies

$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \leqslant 1} |L(f)| \leqslant \|g\|_1 < \infty.$$

On the other hand, the function $f = \operatorname{sign} g$ belongs $L^\infty(\mathbb{R}^n)$ and thus

$$|L(f)| = \int_{\mathbb{R}^n} g \operatorname{sign} g \, d\mu = \int_{\mathbb{R}^n} |g| \, d\mu = \|g\|_1.$$

This shows that

$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \leqslant 1} |L(f)| \geqslant \|g\|_1,$$

from which it follows that $\|L\| = \|g\|_1$.

$\boxed{p = 1}$ Let $g \in L^\infty(\mathbb{R}^n)$. Assume first that $\mu(\mathbb{R}^n) < \infty$. Let $\varepsilon > 0$ and set

$$A_\varepsilon = \{x \in \mathbb{R}^n : |g(x)| \geqslant \|g\|_\infty - \varepsilon\} \quad \text{and} \quad f = \frac{\chi_{A_\varepsilon} \operatorname{sign} g}{\mu(A_\varepsilon)}.$$

Observe that $0 < \mu(A_\varepsilon) \leqslant \mu(\mathbb{R}^n) < \infty$. Then $f \in L^1(\mathbb{R}^n)$, since

$$\|f\|_1 = \int_{\mathbb{R}^n} \frac{\chi_{A_\varepsilon}}{\mu(A_\varepsilon)} \, d\mu = 1.$$

Thus

$$|L(f)| = \int_{A_\varepsilon} \frac{|g|}{\mu(A_\varepsilon)} \, d\mu \geqslant \|g\|_\infty - \varepsilon.$$

The claim follows from this in the case $\mu(\mathbb{R}^n) < \infty$ by letting $\varepsilon \to 0$.

The case $\mu(\mathbb{R}^n) = \infty$ follows by exhausting $\mathbb{R}^n$ with sets $A_i \subset A_{i+1}$, $\mathbb{R}^n = \bigcup_{i=1}^\infty A_i$ with $\mu(A_i) < \infty$ for every $i = 1, 2, \ldots$. Choosing

$$f = \frac{\chi_{A_{i,\varepsilon}} \operatorname{sign} g}{\mu(A_{i,\varepsilon})}$$

in the argument above, we obtain

$$\|g\|_{L^\infty(A_{i,\varepsilon})} - \varepsilon \leqslant \|L\| \leqslant \|g\|_{L^\infty(\mathbb{R}^n)} \quad \text{for every} \quad i = 1, 2, \ldots.$$

The claim follows from this by first letting $\varepsilon \to 0$ and then $i \to \infty$. □

Then we show that the converse of the previous theorem holds.

Theorem 5.2 (Riesz representation theorem in L^p). Let $1 \leqslant p < \infty$ and assume that μ is a Radon measure. Then for every bounded linear functional $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ there corresponds a unique $g \in L^{p'}(\mathbb{R}^n)$ such that

$$L(f) = \int_{\mathbb{R}^n} fg\,d\mu \quad \text{for every} \quad f \in L^p(\mathbb{R}^n). \tag{5.1}$$

Moreover, $\|L\| = \|g\|_{p'}$.

T H E M O R A L : The dual space of $L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$ is isomorphic to $L^{p'}(\mathbb{R}^n)$.

W A R N I N G : The result does not hold for $p = \infty$, since $L^\infty(\mathbb{R}^n)^*$ is not a subset of $L^1(\mathbb{R}^n)$.

Proof. (1) We assume first that $\mu(\mathbb{R}^n) < \infty$. For any measurable set A, define

$$\nu(A) = L(\chi_A).$$

Since we are assuming that $\mu(\mathbb{R}^n) < \infty$, we have $\chi_A \in L^p(\mathbb{R}^n)$ and so ν is well defined.

(2) ν is countable additive on disjoint measurable sets.

Reason. If A and B are disjoint measurable sets, then $\chi_{A\cup B} = \chi_A + \chi_B$ and thus

$$\nu(A \cup B) = \nu(A) + \nu(B).$$

The continuity of L allows us to extend this to countable additivity. If $A_1, A_2, \dots$ are pairwise disjoint measurable sets, $E = \bigcup_i^\infty A_i$ and $E_k = \bigcup_i^k A_i$, then

$$\|\chi_E - \chi_{E_k}\|_p^p = \int_{\mathbb{R}^n} |\chi_E - \chi_{E_k}|^p\,d\mu = \mu(E \setminus E_k) \to 0 \quad \text{as} \quad k \to \infty,$$

since $E \setminus E_k = \bigcup_{i=k+1}^\infty A_i$ and $\mu(\mathbb{R}^n) < \infty$. It follows that $\chi_{E_k} \to \chi_E$ in $L^p(\mathbb{R}^n)$ and by the continuity of L, $L(\chi_{E_k}) \to L(\chi_E)$. Consequently,

$$\sum_{i=1}^k \nu(A_i) = \nu(E_k) \to \nu(E) \quad \text{as} \quad k \to \infty,$$

■

(3) ν is absolutely continuous with respect to μ.

Reason. If $\mu(A) = 0$, then $\chi_A = 0$ in $L^p(\mathbb{R}^n)$ and since a linear functional maps zero to zero, we have $\nu(A) = L(\chi_A) = 0$. ■

(4) It follows that ν is a Radon measure. By the Radon-Nikodym theorem 4.18, there exists $g \in L^1(\mathbb{R}^n)$ for which

$$\nu(A) = L(\chi_A) = \int_{\mathbb{R}^n} \chi_A g\,d\mu$$

for every μ-measurable set $A \subset \mathbb{R}^n$. This proves (5.1) for the characteristic functions. We still have to show that $g \in L^{p'}(\mathbb{R}^n)$, (5.1) holds for all $f \in L^p(\mathbb{R}^n)$ and $\|L\| = \|g\|_{p'}$.

(5) The representation in (5.1) holds for every $f \in L^\infty(\mathbb{R}^n)$.

Reason. In (4) we showed that (5.1) holds for every characteristic function of a measurable set and consequently it holds for linear combinations of such sets. Thus (5.1) holds for simple functions. Every function $f \in L^\infty(\mathbb{R}^n)$ is a uniform limit of a sequence (f_i) of simple functions. Thus

$$\|f_i - f\|_p = \left(\int_{\mathbb{R}^n} |f_i - f|^p \, d\mu\right)^{1/p} \leqslant \|f_i - f\|_\infty \mu(\mathbb{R}^n) \to 0 \quad \text{as} \quad i \to \infty$$

and

$$|L(f_i) - L(f)| \leqslant \|L\| \|f_i - f\|_p \to 0 \quad \text{as} \quad i \to \infty$$

On the other hand,

$$\begin{aligned}\left|\int_{\mathbb{R}^n} f_i g \, d\mu - \int_{\mathbb{R}^n} f g \, d\mu\right| &\leqslant \int_{\mathbb{R}^n} |f_i - f||g| \, d\mu \\ &\leqslant \|f_i - f\|_\infty \int_{\mathbb{R}^n} |g| \, d\mu \to 0 \quad \text{as} \quad i \to \infty.\end{aligned}$$

Thus

$$L(f) = \lim_{i\to\infty} L(f_i) = \lim_{i\to\infty} \int_{\mathbb{R}^n} f_i g \, d\mu = \int_{\mathbb{R}^n} f g \, d\mu.$$

This shows (5.1) holds for every $f \in L^\infty(\mathbb{R}^n)$. ■

(6) $\|g\|_{p'} \leqslant \|L\|$ and thus $g \in L^{p'}(\mathbb{R}^n)$.

Reason. $1 < p < \infty$ Let

$$A_i = \{x \in \mathbb{R}^n : |g(x)| \leqslant i\} \quad \text{and} \quad f = \chi_{A_i} |g|^{p'-1} \operatorname{sign} g, \quad i = 1, 2, \dots.$$

Then $f \in L^\infty(\mathbb{R}^n)$ and $|f|^p = |g|^{p'}$ on A_i. Thus

$$\int_{A_i} |g|^{p'} \, d\mu = \int_{\mathbb{R}^n} f g \, d\mu = L(f) \leqslant \|L\| \|f\|_p = \|L\| \left(\int_{A_i} |g|^{p'} \, d\mu\right)^{1/p}.$$

This implies

$$\int_{A_i} |g|^{p'} \, d\mu \leqslant \|L\|^{p'} \quad \text{for every} \quad i = 1, 2, \dots,$$

and letting $i \to \infty$, by the monotone convergence theorem, we have $\|g\|_{p'} \leqslant \|L\|$.

$p = 1, p' = \infty$ Let

$$A_i = \left\{x \in \mathbb{R}^n : |g(x)| \geqslant \|L\| + \frac{1}{i}\right\}, \quad i = 1, 2, \dots.$$

Then

$$\begin{aligned}\|L\| \|\chi_{A_i}\|_1 &\geqslant |L(\chi_{A_i} \operatorname{sign} g)| \\ &= \left|\int_{\mathbb{R}^n} \chi_{A_i} g \operatorname{sign} g \, d\mu\right| \geqslant \left(\|L\| + \frac{1}{i}\right) \mu(A_i),\end{aligned}$$

which can happen only if $\mu(A_i) = 0$. Since

$$\mu(\{x \in \mathbb{R}^n : |g(x)| > \|L\|\}) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leqslant \sum_{i=1}^{\infty} \mu(A_i) = 0,$$

we have $|g(x)| \leqslant \|L\|$ for almost every $x \in \mathbb{R}^n$. This implies $\|g\|_\infty \leqslant \|L\|$. ■

(7) Thus $g \in L^{p'}(\mathbb{R}^n)$ and

$$L(f) = \int_{\mathbb{R}^n} f g \, d\mu \tag{5.2}$$

for every $f \in L^\infty(\mathbb{R}^n)$. Both sides of (5.2) are continuous linear functionals on $L^p(\mathbb{R}^n)$ and they coincide on the dense subset $L^\infty(\mathbb{R}^n)$. Consequently, they coincide on the whole of $L^p(\mathbb{R}^n)$. This proves that (5.1) holds for all functions $f \in L^p(\mathbb{R}^n)$.

(8) $\|g\|_{p'} = \|L\|$.

Reason. By (6) we have $\|g\|_{p'} \leqslant \|L\|$. The opposite direction comes from Hölder's inquality, since

$$|L(f)| \leqslant \left| \int_{\mathbb{R}^n} f g \, d\mu \right| \leqslant \|f\|_p \|g\|_{p'}$$

so that

$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \leqslant 1} |L(f)| \leqslant \|g\|_{p'}.$$

■

(9) The proof is now complete in the case $\mu(\mathbb{R}^n) < \infty$. Now we consider the case $\mu(\mathbb{R}^n) = \infty$.

(10) For every σ-finite measure on $\mathbb{R}^n$, there exists a function $w \in L^1(X;\mu)$ such that $0 < w(x) < 1$ for every $x \in \mathbb{R}^n$.

Reason. To say that μ is σ-finite means that $\mathbb{R}^n$ is the union of a countably many measurable sets A_i, $i = 1, 2, \dots$, for which $\mu(A_i) < \infty$. Define $w_i(x) = 0$, if $x \in \mathbb{R}^n \setminus A_i$ and

$$w_i(x) = \frac{2^{-i}}{1 + \mu(A_i)}$$

if $x \in A_i$. Then $w = \sum_{i=1}^\infty w_i$ has the required properties. ■

(11) Define a measure

$$\widetilde{\mu}(A) = \int_A w \, d\mu$$

for μ-measurable sets $A \subset \mathbb{R}^n$. Then $\widetilde{\mu}(\mathbb{R}^n) < \infty$ and $\widetilde{f} \mapsto w^{1/p} \widetilde{f}$ is a linear isometry from $L^p(\mathbb{R}^n; \widetilde{\mu})$ onto $L^p(\mathbb{R}^n; \mu)$, because $w(x) > 0$ for every $x \in \mathbb{R}^n$. Hence

$$\widetilde{L}(\widetilde{f}) = L(w^{1/p} \widetilde{f})$$

defines a bounded linear functional $\widetilde{L} : L^p(\mathbb{R}^n; \widetilde{\mu}) \to L^p(\mathbb{R}^n; \widetilde{\mu})$ with $\|T\| = \|L\|$. The beginning of the proof shows that there exists $\widetilde{g} \in L^{p'}(\mathbb{R}^n; \widetilde{\mu})$ such that

$$\widetilde{L}(\widetilde{f}) = \int_{\mathbb{R}^n} \widetilde{f} \widetilde{g} \, d\widetilde{\mu}$$

for every $\widetilde{f} \in L^p(\mathbb{R}^n;\widetilde{\mu})$. Set $g = w^{1/p'}\widetilde{g}$ for $1 < p < \infty$ and $g = \widetilde{g}$ for $p = 1$. Then

$$\int_{\mathbb{R}^n} |g|^{p'}\,d\mu = \int_{\mathbb{R}^n} |\widetilde{g}|^{p'}\,d\widetilde{\mu} = \|\widetilde{L}\|^{p'} = \|L\|^{p'},$$

if $p > 1$. If $p = 1$, then

$$\|g\|_\infty = \|\widetilde{g}\|_\infty = \|\widetilde{L}\| = \|L\|.$$

Finally,

$$L(f) = \widetilde{L}(w^{-1/p}f) = \int_{\mathbb{R}^n} w^{-1/p} f\widetilde{g}\,d\widetilde{\mu} = \int_{\mathbb{R}^n} fg\,d\mu$$

for every $f \in L^p(\mathbb{R}^n;\mu)$. □

Remarks 5.3:

(1) For $p = p' = 2$ the Riesz representation theorem can be proved using the facts that $L^2(\mathbb{R}^n)$ is a complete space and therefore a Hilbert space, and that bounded linear functionals on a Hilbert space are given by inner products.

(2) It follows that $L^p(\mathbb{R}^n)^{**} = L^p(\mathbb{R}^n)$, and thus $L^p(\mathbb{R}^n)$ is a reflexive space when $1 < p < \infty$.

(3) The result does not hold for $p = 1$ without the σ-finiteness assumption. For $p > 1$, we do not need to assume that μ is σ-finite in fact, although the proof requires some different details.

Remarks 5.4:

(1) It follows that

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}^n} fg\,dx : g \in L^{p'}(\mathbb{R}^n),\ \|g\|_{p'} \leqslant 1\right\}.$$

(2) Since $C_0(\mathbb{R}^n)$ is dense in $L^{p'}(\mathbb{R}^n)$ for $1 < p < \infty$, we have

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}^n} fg\,dx : g \in C_0(\mathbb{R}^n),\ \|g\|_{p'} \leqslant 1\right\}.$$

5.2 Partitions of unity

In this section we briefly discuss partitions of unity which are tools to localize problems in analysis to a neighbourhood of each point. In general, a partition of unity is a collection Φ of continuous functions φ with $0 \leqslant \varphi \leqslant 1$ for every $\varphi \in \Phi$ and $\sum_{\varphi\in\Phi}\varphi(x) = 1$ such that for every point x there is a neighbourhood of x such that only finitely many of the functions are nonzero. This ensures that the sum above has only finitely many nonzero terms at each point and that its partial sums are well defined continuous functions. We shall consider a special case of this.

Theorem 5.5. Let $K \subset \mathbb{R}^n$ be a compact set and U_i, $i = 1,2,\dots,k$, be a collection of finitely many open subsets of $\mathbb{R}^n$ such that $K \subset \bigcup_{i=1}^k U_i$. There exist functions $\varphi_i \in C_0(\mathbb{R}^n)$, $i = 1,\dots,k$, such that

(1) $0 \leqslant \varphi_i \leqslant 1$ for every $i = 1, \dots, k$,

(2) $\operatorname{supp}\varphi_i$ is a compact subset of U_i for every $i = 1, \dots, k$ and

(3) $\sum_{i=1}^k \varphi_i(x) = 1$ for every $x \in K$.

TERMINOLOGY: The collection of functions φ_i is called a partition of unity related to the collection U_i, $i = 1, 2, \dots, k$.

THE MORAL: Partition of unity is a very useful tool to localize functions, since

$$f(x) = f(x)\sum_{i=1}^k \varphi_i(x) = \sum_{i=1}^k f(x)\varphi_i(x) \quad \text{for every} \quad x \in K.$$

Thus every function can be represented as a sum of compactly supported functions.

Proof. For every $x \in K$, there exists a ball $B(x, r_x)$ such that $B(x, 2r_x) \subset U_i$ for some $i = 1, \dots, k$. Since K is compact, there exists a finite subcollection $B(x_i, r_i)$, $i = 1, \dots, m$, such that

$$K \subset \bigcup_{i=1}^m B(x_i, r_i).$$

Let A_i be the union of those closed balls $\overline{B}(x_i, r_i)$ for which $B(x_i, 2r_i) \subset U_i$. Then A_i is a compact subset of U_i and $K \subset \bigcup_{i=1}^k A_i$. As in (3.1), we can obtain a cutoff function $g_i \in C(\mathbb{R}^n)$ such that

(1) $0 \leqslant g_i \leqslant 1$,

(2) $g_i = 1$ in A_i and

(3) $\operatorname{supp} g_i$ is a compact subset of U_i.

Define

$$\begin{aligned} \varphi_1 &= g_1, \\ \varphi_2 &= (1 - g_1)g_2, \\ &\vdots \\ \varphi_k &= (1 - g_1)\dots(1 - g_{k-1})g_k. \end{aligned}$$

Clearly $0 \leqslant \varphi_i \leqslant 1$ and $\operatorname{supp}\varphi_i \subset U_i$ for every $i = 1, 2 \dots, k$. By induction

$$\sum_{i=1}^k \varphi_i = 1 - (1 - g_1)\dots(1 - g_k).$$

In addition,

$$0 \leqslant \sum_{i=1}^k \varphi_i \leqslant 1 \quad \text{and} \quad \sum_{i=1}^k \varphi_i = 1 \quad \text{on} \quad K,$$

since if $x \in K$, then $x \in A_i$ for some $i = 1, \dots, k$ and thus $1 - g_i(x) = 0$. □

Remark 5.6. The proof above applies in any locally compact Hausdorff space.

Remark 5.7. Let us briefly discuss another approach to the partition of unity.

(1) Assume that the balls $B(x_i,r_i)$, $i=1,2,\dots$, have the property that even the enlarged balls $B(x_i,2r_i)$, $i=1,2,\dots$, are of bounded overlap, see the Besicovitch covering theorem (Theorem (4.2)) for the bounded overlap property. For every $i=1,2,\dots$, there is a continuous function g_i such that $0\leqslant g_i\leqslant 1$, $g_i=1$ on $B(x_i,r_i)$ and $\operatorname{supp} g_i=\overline{B}(x_i,2r_i)$. Define

$$\varphi_i(x)=\frac{g_i(x)}{\sum_{j=1}^{\infty} g_j(x)},\quad i=1,2,\dots.$$

Observe that the sum is only over finitely many terms at a given point. Then

$$\sum_{i=1}^{\infty}\varphi_i=1\quad\text{in}\quad\bigcup_{i=1}^{\infty}B(x_i,2r_i).$$

(2) This method applies to other sets than balls as well. Let U be a nonempty open set in $\mathbb{R}^n$. Denote $U_0=\emptyset$ and

$$U_i=\left\{x\in U:\operatorname{dist}(x,\partial U)>\frac{1}{i}\right\}\cap B(0,i),\quad i=1,2,\dots.$$

Then $U=\bigcup_{i=1}^{\infty}U_i$ and $\overline{U_i}$ is a compact subset of U_{i+1} for every $i=1,2,\dots$.
CLAIM: There exists $\varphi_i\in C_0^{\infty}(U_{i+2}\setminus\overline{U_{i-1}})$, $0\leqslant\varphi_i\leqslant 1$ for every $i=1,2,\dots$ and

$$\sum_{i=1}^{\infty}\varphi_i=1\quad\text{in}\quad U.$$

Reason. There exists $g_i\in C_0^{\infty}(U_{i+2}\setminus\overline{U_{i-1}})$ with $0\leqslant g_i\leqslant 1$ and $g_i=1$ in $\overline{U_{i+1}}\setminus U_i$ for every $i=1,2,\dots$. Define $\varphi_i:U\to\mathbb{R}$,

$$\varphi_i(x)=\frac{g_i(x)}{\sum_{j=1}^{\infty} g_j(x)},\quad i=1,2,\dots.$$

■

Let $f\in L^p(U)$ with $1\leqslant p<\infty$ and assume that ϕ_ε be an approximation of the identity as in Defintion 3.14. Then $\varphi_i f$ has a compact support and $\operatorname{supp}(\varphi_i f)\subset U_{i+2}\setminus\overline{U_{i-1}}$. Fix $\varepsilon>0$. Choose $\varepsilon_i>0$ so small that

$$\operatorname{supp}(\phi_{\varepsilon_i}*(\varphi_i f))\subset U_{i+2}\setminus\overline{U_{i-1}}$$

and

$$\|\phi_{\varepsilon_i}*(\varphi_i f)-\varphi_i f\|_{L^p(U)}<\frac{\varepsilon}{2^i},\quad i=1,2,\dots.$$

Define

$$g=\sum_{i=1}^{\infty}\phi_{\varepsilon_i}*(\varphi_i f).$$

This function belongs to $C^{\infty}(U)$, since in a neighbourhood of any point $x\in U$, there are only finitely many nonzero terms in the sum. Moreover,

$$\begin{aligned}\|f-g\|_{L^p(U)}&=\left\|\sum_{i=1}^{\infty}\phi_{\varepsilon_i}*(\varphi_i f)-\sum_{i=1}^{\infty}\varphi_i f\right\|_{L^p(U)}\\&\leqslant\sum_{i=1}^{\infty}\left\|\phi_{\varepsilon_i}*(\varphi_i f)-\sum_{i=1}^{\infty}\varphi_i f\right\|_{L^p(U)}\leqslant\sum_{i=1}^{\infty}\frac{\varepsilon}{2^i}=\varepsilon.\end{aligned}$$

This shows that $C^\infty(U)$ is dense in $L^p(U)$ for $1 \leqslant p < \infty$ for an arbitrary open subset U of $\mathbb{R}^n$. It can be shown that even $C_0^\infty(U)$ is dense in $L^p(U)$ for $1 \leqslant p < \infty$ for an arbitrary open subset U of $\mathbb{R}^n$ (exercise).

5.3 The Riesz representation theorem for Radon measures

We denote the space of continuous functions $f : \mathbb{R}^n \to \mathbb{R}^m$ by $C(\mathbb{R}^n;\mathbb{R}^m)$, where $n, m = 1, 2, \ldots$. The support of such a function f is

$$\operatorname{supp} f = \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}$$

and

$$C_0(\mathbb{R}^n;\mathbb{R}^m) = \left\{ f \in C(\mathbb{R}^n;\mathbb{R}^m) : \operatorname{supp} f \text{ is a compact subset of } \mathbb{R}^n \right\}$$

is the space of compactly supported continuous functions $f : \mathbb{R}^n \to \mathbb{R}^m$.

Assume that μ is a Radon measure on $\mathbb{R}^n$ and let $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ be a μ-measurable function such that $|\sigma(x)| = 1$ for μ-almost every $x \in \mathbb{R}^n$. Define $L : C_0(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}$,

$$L(f) = \int_{\mathbb{R}^n} f \cdot \sigma \, d\mu$$

for every $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$. Then

$$L(f+g) = L(f) + L(g) \quad \text{and} \quad L(af) = aL(f), \quad a \in \mathbb{R},$$

so that L is a linear functional. Let $K \subset \mathbb{R}^n$ be a compact set and assume that $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ with $\operatorname{supp} f \subset K$ and $|f(x)| \leqslant 1$ for every $x \in \mathbb{R}^n$. Then

$$\begin{aligned} |L(f)| &= \left| \int_{\mathbb{R}^n} f \cdot \sigma \, d\mu \right| \leqslant \int_{\mathbb{R}^n} |f \cdot \sigma| \, d\mu \leqslant \int_{\mathbb{R}^n} |f||\sigma| \, d\mu \\ &= \int_{\mathbb{R}^n} |f| \, d\mu \leqslant \mu(\operatorname{supp} f) \leqslant \mu(K) < \infty. \end{aligned}$$

Thus

$$\sup \left\{ |L(f)| : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset K \right\} < \infty.$$

This is the norm of the linear functional L over the class of functions $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ with $\operatorname{supp} f \subset K$. Thus this functional is locally bounded.

THE MORAL: The integral with respect to a Radon measure defines a locally bounded linear functional $L : C_0(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}$ as above.

The next theorem shows that the converse holds as well.

Theorem 5.8 (Riesz representation theorem). Assume that $L : C_0(\mathbb{R}^n;\mathbb{R}^m) \to \mathbb{R}$ is a linear functional satisfying

$$\sup \left\{ L(f) : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset K \right\} < \infty \tag{5.3}$$

for every compact set $K \subset \mathbb{R}^n$. Then there exists a Radon measure μ on $\mathbb{R}^n$ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that $|\sigma(x)| = 1$ for μ-almost every $x \in \mathbb{R}^n$ and

$$L(f) = \int_{\mathbb{R}^n} f \cdot \sigma \, d\mu$$

for every $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$.

THE MORAL: This means that a locally bounded linear functional on $C_0(\mathbb{R}^n;\mathbb{R}^m)$ can be characterized as an integral with respect to a Radon measure. The role of σ is just to assign a sign so that the measure μ is nonnegative.

Example 5.9. Let $L : C_0(\mathbb{R}^n) \to \mathbb{R}$, $L(f) = f(x_0)$ be the evaluation map for a fixed $x_0 \in \mathbb{R}^n$. Let $K \subset \mathbb{R}^n$ be a compact set, $f \in C_0(\mathbb{R}^n)$ with $|f| \leqslant 1$ and $\operatorname{supp} f \subset K$. Then

$$L(f) = f(x_0) \leqslant 1 < \infty.$$

This functional is positive in the sense that $L(f) \geqslant 0$ whenever $f \geqslant 0$. By the Riesz representation theorem, there exists a Radon measure μ on $\mathbb{R}^n$ such that

$$L(f) = \int_{\mathbb{R}^n} f \, d\mu$$

for every $f \in C_0(\mathbb{R}^n)$. It follows that for the evaluation map, the measure μ is equal to Dirac's measure δ_{x_0} concentrated at x_0.

Definition 5.10. The variation measure μ is defined as

$$\mu(U) = \sup\left\{L(f) : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset U\right\}$$

for every open set $U \subset \mathbb{R}^n$.

Now we are ready for the proof od Theorem 5.8.

Proof. (1) Define the variation measure for open sets as above and set

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

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

(2) μ is an outer measure.

Reason. Let U and U_i, $i = 1,2,\dots$, be open subsets of $\mathbb{R}^n$ such that $U \subset \bigcup_{i=1}^{\infty} U_i$. Choose $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ such that $|f| \leqslant 1$ and $\operatorname{supp} f \subset U$. Since $\operatorname{supp} f$ is a compact set and the collection of sets U_i, $i = 1,2,\dots$, is an open covering of K, there are finitely many sets U_i, $i = 1,\dots,k$, such that $\operatorname{supp} f \subset \bigcup_{i=1}^{k} U_i$.

Let φ_i, $i = 1,\dots,k$, be a partition of unity related to the collection U_i, $i = 1,\dots,k$, such that $0 \leqslant \varphi_i \leqslant 1$, $\operatorname{supp} \varphi_i \subset U_i$ for every $i = 1,\dots,k$, and and

$$\sum_{i=1}^{k} \varphi_i(x) = 1 \quad \text{for every} \quad x \in \operatorname{supp} f.$$

Then

$$f(x) = f(x)\sum_{i=1}^{k}\varphi_i(x) = \sum_{i=1}^{k} f(x)\varphi_i(x) \quad \text{for every} \quad x \in \mathbb{R}^n,$$

$\operatorname{supp}(\varphi_i f) \subset U_i$ and $0 \leqslant \varphi_i f \leqslant 1$ for every $i = 1,2,\dots$. Thus

$$|L(g)| = \left|L\left(\sum_{i=1}^{k} f\varphi_i\right)\right| = \left|\sum_{i=1}^{k} L(f\varphi_i)\right| \leqslant \sum_{i=1}^{k} |L(f\varphi_i)| \leqslant \sum_{i=1}^{\infty} \mu(U_i).$$

By taking the supremum over such functions f, we have

$$\mu(U) \leqslant \sum_{i=1}^{\infty} \mu(U_i).$$

Then let A_i, $i = 1,2,\dots$, be arbitrary sets with $A \subset \bigcup_{i=1}^{\infty} A_i$. Fix $\varepsilon > 0$. For every $i = 1,2,\dots$, choose an open set U_i such that $A_i \subset U_i$ and

$$\mu(U_i) \leqslant \mu(A_i) + \frac{\varepsilon}{2^i}.$$

Then

$$\mu(A) \leqslant \mu\left(\bigcup_{i=1}^{\infty} U_i\right) \leqslant \sum_{i=1}^{\infty} \mu(U_i) \leqslant \sum_{i=1}^{\infty}\left(\mu(A_i) + \frac{\varepsilon}{2^i}\right) = \sum_{i=1}^{\infty} \mu(A_i) + \varepsilon.$$

■

(3) μ is a Radon measure.

Reason. Assume first that U_1 and U_2 are open sets with $\operatorname{dist}(U_1,U_2) > 0$. Let $f_i \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ such that $0 \leqslant f_i \leqslant 1$ and $\operatorname{supp} f_i \subset U_i$, $i = 1,2$. Then $f_1 + f_2 \in C_0(\mathbb{R}^n;\mathbb{R}^m)$, $0 \leqslant f_1 + f_2 \leqslant 1$ and $\operatorname{supp}(f_1 + f_2) \subset U_1 \cup U_2$, so that

$$L(f_1) + L(f_2) = L(f_1 + f_2) \leqslant \mu(U_1 \cup U_2).$$

By taking the supremum over all admissible functions f_1 and f_2, we obtain

$$\mu(U_1) + \mu(U_2) \leqslant \mu(U_1 \cup U_2).$$

On the other hand, by (1) we have $\mu(U_1 + U_2) \leqslant \mu(U_1) + \mu(U_2)$. Thus

$$\mu(U_1) + \mu(U_2) = \mu(U_1 + U_2).$$

Assume then that A_1 and A_2 are arbitrary sets with $\operatorname{dist}(A_1,A_2) > 0$. Let $\varepsilon > 0$ and choose an open set $U \subset \mathbb{R}^n$ such that $A_1 \cup A_2 \subset U$ and

$$\mu(U) \leqslant \mu(A_1 \cup A_2) + \varepsilon.$$

Take open sets $U_i \subset \mathbb{R}^n$ such that $A_i \subset U_i$ and $\operatorname{dist}(U_1,U_2) > 0$. For example, we may take

$$U_i = \left\{x \in \mathbb{R}^n : \operatorname{dist}(x,A_i) < \frac{1}{3}\operatorname{dist}(A_1,A_2)\right\}, \quad i = 1,2.$$

Then $A_i \subset U_i \cap U$ and $\operatorname{dist}(U_1 \cap U, U_2 \cap U) > 0$. Thus

$$\begin{aligned}\mu(A_1) + \mu(A_2) &\leqslant \mu(U_1 \cap U) + \mu(U_2 \cap U)\\ &= \mu(U \cap (U_1 \cap U_2)) \leqslant \mu(A_1 \cup A_2) + \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$, we have $\mu(A_1 \cup A_2) \leqslant \mu(A_1) + \mu(A_2)$. Again, the reverse inequality follows by subadditivity, so that

$$\mu(A_1 \cup A_2) \leqslant \mu(A_1) + \mu(A_2).$$

This shows that μ is a metric outer measure and consequently it is a Borel measure.

To see that μ is Borel regular, let A be an arbitrary subset of $\mathbb{R}^n$. Then there exist open sets U_i, $i = 1, 2, \dots$, such that $A \subset U_i$ and

$$\mu(U_i) < \mu(A) + \frac{1}{i}, \quad i = 1, 2, \dots.$$

Thus

$$\mu(A) \leqslant \mu\left(\bigcap_{i=1}^{\infty} U_i\right) \leqslant \mu(U_i) < \mu(A) + \frac{1}{i}$$

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

$$\mu(A) = \mu\left(\bigcap_{i=1}^{\infty} U_i\right).$$

Finally, we show that $\mu(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$. It is enough to show that $\mu(B(x,r)) < \infty$ for every ball $B(x,r) \subset \mathbb{R}^n$. This is clear, since (5.3) gives

$$\begin{aligned} \mu(B(x,r)) &= \sup\left\{L(f) : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset B(x,r)\right\} \\ &\leqslant \sup\left\{L(f) : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset \overline{B}(x,r)\right\} < \infty. \end{aligned}$$

Thus μ satisfies all conditions in the definition of the Radon measure. ■

(4) Denote $C_0^+(\mathbb{R}^n) = \{f \in C_0(\mathbb{R}^n) : f \geqslant 0\}$ and for every $f \in C_0^+(\mathbb{R}^n)$ define

$$\nu(f) = \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m), |g| \leqslant f\right\}.$$

Observe that if $f_1, f_2 \in C_0^+(\mathbb{R}^n)$ and $f_1 \leqslant f_2$, then $\nu(f_1) \leqslant \nu(f_2)$. Moreover, $\nu(af) = a\nu(f)$ for every $a \in \mathbb{R}$ and $f \in C_0^+(\mathbb{R}^n)$.

(5) $\nu(f_1 + f_2) = \nu(f_1) + \nu(f_2)$ for every $f_1, f_2 \in C_0^+(\mathbb{R}^n)$.

Reason. If $g_1, g_2 \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ such that $|g_1| \leqslant f_1$ and $|g_2| \leqslant f_2$, then $|g_1 + g_2| \leqslant |g_1| + |g_2| \leqslant f_1 + f_2$. In addition, we may assume that $L(g_1) \geqslant 0$ and $L(g_2) \geqslant 0$. Thus

$$|L(g_1)| + |L(g_2)| = L(g_1) + L(g_2) = L(g_1 + g_2) \leqslant |L(g_1 + g_2)| \leqslant \nu(f_1 + f_2).$$

By taking suprema over $g_1 \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ and $g_2 \in C_0(\mathbb{R}^n;\mathbb{R}^m)$, we have

$$\nu(f_1) + \nu(f_2) \leqslant \nu(f_1 + f_2).$$

To prove the reverse inequality, let $g \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ with $|g| \leqslant f_1 + f_2$. For $i = 1, 2$, set

$$g_i = \begin{cases} \dfrac{f_i g}{f_1 + f_2}, & \text{if} \quad f_1 + f_2 > 0, \\ 0, & \text{if} \quad f_1 + f_2 = 0, \end{cases}$$

Then $g_1, g_2 \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ and $g = g_1 + g_2$. Moreover, $|g_1| \leqslant f_1$ and $|g_2| \leqslant f_2$, so that

$$|L(g)| \leqslant |L(g_1)| + |L(g_2)| \leqslant \nu(f_1) + \nu(f_2),$$

from which it follows that $\nu(f_1 + f_2) \leqslant \nu(f_1) + \nu(f_2)$. ∎

(6) $\nu(f) = \int_{\mathbb{R}^n} f\,d\mu$ for every $f \in C_0^+(\mathbb{R}^n)$.

Reason. Let $\varepsilon > 0$. Choose $0 = t_0 < t_1 < \cdots < t_k$ such that

$$t_k = 2\|f\|_\infty, \quad 0 < t_i - t_{i-1} < \varepsilon \quad \text{and} \quad \mu(f^{-1}\{t_i\}) = 0 \quad \text{for every} \quad i = 1, \dots, k.$$

Set $U_i = f^{-1}((t_{i-1}, t_i))$, then U_i is open and $\mu(U_i) < \infty$ for every $i = 1, \dots, k$. By approximation properties of measurable sets with respect to a Radon measure, there exist compact sets $K_i \subset U_i$ such that

$$\mu(U_i \setminus K_i) < \frac{\varepsilon}{k} \quad i = 1, \dots, k.$$

Futhermore, there exist functions $g_i \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ with $|g_i| \leqslant 1$, $\operatorname{supp} g_i \subset U_i$ such that

$$|L(g_i)| \geqslant \mu(U_i) - \frac{\varepsilon}{k}, \quad i = 1, \dots, k.$$

Note also that there exit functions $h_i \in C_0^+(\mathbb{R}^n)$ such that $\operatorname{supp} h_i \subset U_i$, $0 \leqslant h_i \leqslant 1$ and $h_i = 1$ in the compact set $K_i \cup \operatorname{supp} g_i$. Then

$$\nu(h_i) \geqslant |L(g_i)| \geqslant \mu(U_i) - \frac{\varepsilon}{k}, \quad i = 1, \dots, k,$$

and

$$\begin{aligned}\nu(h_i) &= \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\, |g| \leqslant h_i\right\} \\ &\leqslant \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\, |g| \leqslant 1, \operatorname{supp} g \subset U_i\right\} = \mu(U_i).\end{aligned}$$

Thus

$$\mu(U_i) - \frac{\varepsilon}{k} \leqslant \nu(h_i) \leqslant \mu(U_i), \quad i = 1, \dots, k.$$

Define

$$A = \left\{x \in \mathbb{R}^n : f(x)\left(1 - \sum_{i=1}^k h_i(x)\right) > 0\right\}.$$

Then A is an open set. We have

$$\begin{aligned}\nu\left(f - f\sum_{i=1}^k h_i\right) &= \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\, |g| \leqslant f - f\sum_{i=1}^k h_i\right\} \\ &\leqslant \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\, |g| \leqslant \|f\|_\infty \chi_A\right\} \\ &= \|f\|_\infty \sup\left\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\, |g| \leqslant \chi_A\right\} \\ &= \|f\|_\infty \mu(A) = \|f\|_\infty \mu\left(\bigcup_{i=1}^k (U_i \setminus \{h_i = 1\})\right) \\ &= \|f\|_\infty \sum_{i=1}^k \mu(U_i \setminus K_i) \leqslant \varepsilon \|f\|_\infty.\end{aligned}$$

Thus

$$\begin{aligned}\nu(f) &= \nu\left(f - f\sum_{i=1}^{k} h_i\right) + \nu\left(f\sum_{i=1}^{k} h_i\right)\\ &\leqslant \varepsilon\|f\|_\infty + \sum_{i=1}^{k}\nu(fh_i)\\ &\leqslant \varepsilon\|f\|_\infty + \sum_{i=1}^{k} t_i\mu(U_i)\end{aligned}$$

and

$$\nu(f) \geqslant \sum_{i=1}^{k}\nu(fh_i) \geqslant \sum_{i=1}^{k} t_{i-1}\left(\mu(U_i) - \frac{\varepsilon}{k}\right) \geqslant \sum_{i=1}^{k} t_{i-1}\mu(U_i) - t_k\varepsilon.$$

Since

$$\sum_{i=1}^{k} t_{i-1}\mu(U_i) \leqslant \int_{\mathbb{R}^n} f\,d\mu \leqslant \sum_{i=1}^{k} t_i\mu(U_i),$$

we have

$$\begin{aligned}\left|\nu(f) - \int_{\mathbb{R}^n} f\,d\mu\right| &\leqslant \sum_{i=1}^{k}(t_i - t_{i-1})\mu(U_i) + \varepsilon\|f\|_\infty + \varepsilon t_k\\ &\leqslant \varepsilon\mu(\operatorname{supp} f) + 3\varepsilon\|f\|_\infty.\end{aligned}$$

■

(7) There exists a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$L(f) = \int_{\mathbb{R}^n} f\cdot\sigma\,d\mu$$

for every $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$.

Reason. Fix $e \in \mathbb{R}^n$ with $|e| = 1$. Define $\nu_e(f) = L(fe)$ for every $f \in C_0(\mathbb{R}^n)$. Then ν_e is linear and

$$\begin{aligned}|\nu_e(f)| &= |L(fe)| = \sup\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m),\ |g| \leqslant |f|\}\\ &= \nu(|f|) = \int_{\mathbb{R}^n} |f|\,d\mu.\end{aligned}$$

Thus we can extend ν_e to a bounded linear functional on $L^1(\mathbb{R}^n;\mu)$. Hence there exists $\sigma_e \in L^\infty(\mathbb{R}^n;\mu)$ such that

$$\lambda_e(f) = \int_{R^n} f\sigma_e\,d\mu, \quad f \in C_0(\mathbb{R}^n).$$

Let $e_1,\ldots,e_m$ be the standard basis for $\mathbb{R}^m$ and define $\sigma = \sum_{i=1}^m \sigma_{e_i} e_i$. For $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$, we have

$$L(f) = \sum_{i=1}^{m} L((f\cdot e_i)e_i) = \sum_{i=1}^{m}\int_{\mathbb{R}^n}(f\cdot e_i)e_i\,d\mu = \int_{\mathbb{R}^n} f\cdot\sigma\,d\mu.$$

■

(8) $|\sigma(x)| = 1$ for μ-almost every $x \in \mathbb{R}^n$.

Reason. Let $U \subset \mathbb{R}^n$ be a open set with $\mu(U) < \infty$. By definition

$$\mu(U) = \sup\left\{\int_{\mathbb{R}^n} f\cdot\sigma\,d\mu : f \in C_0(\mathbb{R}^n;\mathbb{R}^m), |f| \leqslant 1, \operatorname{supp} f \subset U\right\}.$$

Let $f_i \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ such that $|f_i| \leqslant 1$, $\operatorname{supp} f_i \subset U$ and $f_i\cdot\sigma \to |\sigma|$ μ-almost everywhere. Thus

$$\int_U |\sigma|\,d\mu = \lim_{i\to\infty}\int_{\mathbb{R}^n} f_i\cdot\sigma\,d\mu \leqslant \mu(U).$$

On the other hand, if $f \in C_0(\mathbb{R}^n;\mathbb{R}^m)$ is such that $|f| \leqslant 1$, $\operatorname{supp} f \subset U$, then

$$\int_{\mathbb{R}^n} f\cdot\sigma\,d\mu \leqslant \int_U |\sigma|\,d\mu$$

and consequently

$$\mu(U) = \int_U |\sigma|\,d\mu.$$

Thus

$$\mu(U) \leqslant \int_U |\sigma|\,d\mu \quad \text{for every open} \quad U \subset \mathbb{R}^n.$$

This implies $|\sigma| = 1$ for μ-almost everywhere. ■

Next we prove the Riesz representation theorem for positive linear functionals on $C_0(\mathbb{R}^n)$.

Theorem 5.11. Assume that $L : C_0(\mathbb{R}^n) \to \mathbb{R}$ is a positive linear functional, that is, $L(f) \geqslant 0$ for every $f \in C_0(\mathbb{R}^n)$ with $f \geqslant 0$. Then there exists a Radon measure μ on $\mathbb{R}^n$ such that

$$L(f) = \int_{\mathbb{R}^n} f\,d\mu$$

for every $f \in C_0(\mathbb{R}^n)$.

Remarks 5.12:

(1) If $f, g \in C_0(\mathbb{R}^n)$ and $f \geqslant g$, then $T(f) - T(g) = T(f-g) \geqslant 0$ and thus $T(f) \geqslant T(g)$.

(2) Positive linear functionals on $C_0(\mathbb{R}^n)$ are not necessarily bounded, but they are locally bounded, as we shall see in the proof below.

Proof. Let K be a compact subset of $\mathbb{R}^n$ and choose $\varphi \in C_0^\infty(\mathbb{R}^n)$ such that $\varphi = 1$ on K and $0 \leqslant \varphi \leqslant 1$ (exercise). Then for every $f \in C_0^\infty(\mathbb{R}^n)$ with $\operatorname{supp} f \subset K$, set

$$g = \|f\|_\infty \varphi - f \geqslant 0.$$

Thus

$$0 \leqslant L(g) = \|f\|_\infty L(\varphi) - L(f)$$

which implies

$$L(f) \leqslant c\|f\|_\infty$$

with $c = L(\varphi)$. The mapping L can be thus extended to a linear mapping from $C_0(\mathbb{R}^n)$ to $\mathbb{R}$, which satisfies the assumptions in the Riesz representation theorem (exercise). Hence there exists a Radon measure μ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}$ such that $|\sigma(x)| = 1$ for μ-almost every $x \in \mathbb{R}^n$ and

$$L(f) = \int_{\mathbb{R}^n} f \sigma \, d\mu$$

for every $f \in C_0(\mathbb{R}^n)$. Then $\sigma(x) = \pm 1$ for μ-almost every $x \in \mathbb{R}^n$ and positivity of the operator implies $\sigma(x) = 1$ for μ-almost every $x \in \mathbb{R}^n$ (exercise). □

Remark 5.13. The Riesz representation theorem holds in a much more general context: The underlying space can be any locally compact Hausdorff space X instead of $\mathbb{R}^n$.

5.4 Weak convergence and compactness of Radon measures

Let us recall the notion of weak convergence from functional analysis. Let X be a normed space. A sequence (x_i) in X is said to be weakly converging, if there exists an element $x \in X$ such that

$$\lim_{i\to\infty} x^*(x_i) = x^*(x) \quad \text{for every} \quad x^* \in X^*.$$

A sequence (x_i^*) in the dual space X^* is said to be weakly (weak star) converging, if there exists an element $x^* \in X^*$ such that

$$\lim_{i\to\infty} x_i^*(x) = x^*(x) \quad \text{for every} \quad x \in X.$$

By the Riesz representation theorem, every bounded linear functional on $C_0(\mathbb{R}^n)$ is an integral with respect to a Radon measure. This gives a motivation for the following definition.

Definition 5.14. The sequence (μ_i) of Radon measures μ_i, $i = 1, 2, \ldots$, converges weakly to the Radon measure μ, if

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} f \, d\mu_i = \int_{\mathbb{R}^n} f \, d\mu \quad \text{for every} \quad f \in C_0(\mathbb{R}^n).$$

In this case we write $\mu_i \rightharpoonup \mu$ as $i \to \infty$.

Examples 5.15:

(1) Let δ_i be Dirac's measure at $i = 1, 2, \ldots$ on $\mathbb{R}$. Then $\delta_i \rightharpoonup 0$ as $i \to \infty$.

(2) Let

$$\mu_i = \frac{1}{i}\left(\delta_{\frac{1}{i}} + \delta_{\frac{2}{i}} + \cdots + \delta_{\frac{i}{i}}\right) \quad i = 1, 2, \ldots.$$

Then for every $f \in C_0(\mathbb{R})$, we have

$$\int_{\mathbb{R}} f \, d\mu_i = \sum_{j=1}^{i} \frac{1}{i} f\left(\frac{j}{i}\right) \to \int_0^1 f(x) \, dx,$$

since these are Riemann sums in $[0,1]$. Thus $\mu_i \rightharpoonup m^1\lfloor_{[0,1]}$ as $i \to \infty$

Lemma 5.16. Assume that μ_i, $i = 1,2,\dots$, are Radon measures on $\mathbb{R}^n$ with $\mu_i \rightharpoonup \mu$ as $i \to \infty$. Then the following claims are true:

(1) $\limsup_{i\to\infty} \mu_i(K) \leqslant \mu(K)$ for every compact set $K \subset \mathbb{R}^n$ and
(2) $\mu(U) \leqslant \liminf_{i\to\infty} \mu_i(U)$ for every open set $U \subset \mathbb{R}^n$.

Proof. $\boxed{(1)}$ Let $K \subset \mathbb{R}^n$ be compact and choose open U such that $K \subset U$. Choose $f \in C_0(\mathbb{R}^n)$ such that $0 \leqslant f \leqslant 1$, $\operatorname{supp} f \subset U$ and $f = 1$ on K. Then

$$\mu(U) \geqslant \int_{\mathbb{R}^n} f\,d\mu = \lim_{i\to\infty} \int_{\mathbb{R}^n} f\,d\mu_i \geqslant \limsup_{i\to\infty} \mu_i(K).$$

Taking infimum over all open sets $U \supset K$, we have

$$\limsup_{i\to\infty} \mu_i(K) \leqslant \inf\{\mu(U) : U \supset K, U \text{ open}\} = \mu(K).$$

$\boxed{(2)}$ Let U be open and $K \subset U$ compact. Assume that the function f is as above. Then

$$\mu(K) \leqslant \int_{\mathbb{R}^n} f\,d\mu = \lim_{i\to\infty} \int_{\mathbb{R}^n} f\,d\mu_i \leqslant \liminf_{i\to\infty} \mu_i(U).$$

Thus

$$\mu(U) = \sup\{\mu(K) : K \subset U, K \text{ compact}\} \leqslant \liminf_{i\to\infty} \mu_i(U).$$

□

Next we prove a very useful weak compactness result for Radon measures.

Theorem 5.17. Let (μ_i), $i = 1,2,\dots$, be a sequence of Radon measures on $\mathbb{R}^n$ with

$$\sup_i \mu_i(K) < \infty$$

for every compact set $K \subset \mathbb{R}^n$. Then there is a subsequence (μ_{i_j}), $j = 1,2,\dots$, and a Radon measure μ such that $\mu_i \rightharpoonup \mu$ as $i \to \infty$.

Proof. $\boxed{(1)}$ Assume first that $M = \sup_i \mu_i(\mathbb{R}^n) < \infty$.

$\boxed{(2)}$ Let $\{f_j\}_{j=1}^{\infty}$ be a countable dense subset of $C_0(\mathbb{R}^n)$ with respect to $\|\cdot\|_\infty$ norm (exercise). Assumption (1) implies that

$$\sup_i \int_{\mathbb{R}^n} f_1\,d\mu_i \leqslant \|f_1\|_\infty \sup_i \mu_i(\mathbb{R}^n) = M\|f_1\|_\infty < \infty.$$

This shows that $\int_{\mathbb{R}^n} f_1\,d\mu_i$ is a bounded sequence in $\mathbb{R}$ and thus it has a converging subsequence. Hence there exists a subsequence (μ_i^1) of (μ_i) and $a_1 \in \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} f_1\,d\mu_i^1 \to a_1 \quad \text{as} \quad i \to \infty.$$

Inductively, we choose a subsequence (μ_i^j) of (μ_i^{j-1}) and $a_j \in \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} f_k\,d\mu_i^k \to a_k \quad \text{as} \quad i \to \infty.$$

Then the diagonal sequence (μ_i^i) satisfies

$$\lim_{i\to\infty}\int_{\mathbb{R}^n} f_k\,d\mu_i^i = a_k \quad \text{for every} \quad k=1,2,\dots.$$

Let S be the vector space spanned by f_k, $k=1,2,\dots$, that is,

$$S=\left\{g=\sum_{k=1}^{m}\lambda_k f_k : \lambda_i\in\mathbb{R},\, m\in\mathbb{N}\right\}.$$

Define

$$L(g)=\sum_{k=1}^{m}\lambda_k a_k, \quad \text{when} \quad g=\sum_{k=1}^{m}\lambda_k f_k.$$

Then

$$\begin{aligned}L(g)&=\sum_{k=1}^{m}\lambda_k a_k=\sum_{k=1}^{m}\lim_{i\to\infty}\int_{\mathbb{R}^n}\lambda_k f_k\,d\mu_i^i\\&=\lim_{i\to\infty}\int_{\mathbb{R}^n}\sum_{k=1}^{m}\lambda_k f_k\,d\mu_i^i=\lim_{i\to\infty}\int_{\mathbb{R}^n} g\,d\mu_i^i\end{aligned}$$

for every $g\in S$. Thus L is a positive linear functional on S. Moreover,

$$|L(g)|=\lim_{i\to\infty}\left|\int_{\mathbb{R}^n} g\,d\mu_i^i\right|\leqslant\sup_i(\|g\|_\infty\mu_i^i(\mathbb{R}^n))\leqslant M\|g\|_\infty. \tag{5.4}$$

This shows that L is a bounded functional on S.

(3) The functional L on S can be uniquely extended to a bounded linear functional on $C_0(\mathbb{R}^n)$.

Reason. Let $f\in C_0(\mathbb{R}^n)$. Since $\{f_j\}_{j=1}^\infty$ is dense in $C_0(\mathbb{R}^n)$, there exists a sequence (f_j) such that $\|f_j-f\|_\infty\to 0$, that is, $f_j\to f$ uniformly in $\mathbb{R}^n$ as $j\to\infty$. Define

$$L(f)=\lim_{j\to\infty}L(f_j).$$

It follows from (5.4) that L is a well defined functional in $C_0(\mathbb{R}^n)$ and that (5.4) holds for every $f\in C_0(\mathbb{R}^n)$. ■

(4) According to the Riesz representation theorem 5.8 there exists a Radon measure μ on $\mathbb{R}^n$ such that

$$L(f)=\int_{\mathbb{R}^n} f\,d\mu \quad \text{for every} \quad f\in C_0(\mathbb{R}^n).$$

Observe, that L is a positive linear functional on $C_0(\mathbb{R}^n)$.

Reason. If $f\in C_0(\mathbb{R}^n)$ with $f\geqslant 0$, and $\|f_j-f\|_\infty\to 0$ with $f_j\in S$, then

$$\liminf_{j\to\infty}(\min_{\mathbb{R}^n} f_j)\geqslant 0$$

and hence

$$L(f)=\lim_{j\to\infty}L(f_j)=\lim_{j\to\infty}\lim_{i\to\infty}\int_{\mathbb{R}^n} f_j\,d\mu_i^i\geqslant 0,$$

since

$$\int_{\mathbb{R}^n} f_j\,d\mu_i^i\geqslant M\min(0,\min_{\mathbb{R}^n} f_j).$$

■

(5) $\mu_i \rightharpoonup \mu$ as $i \to \infty$.

Reason. Let $\varepsilon > 0$ and $f \in C_0(\mathbb{R}^n)$. Choose $g \in S$ such that

$$\|f - g\|_\infty < \frac{\varepsilon}{2M}.$$

Then for large enough i, we have

$$\begin{aligned}\left|L(f) - \int_{\mathbb{R}^n} f\,d\mu_i^i\right| &\leqslant |L(f-g)| + \left|L(g) - \int_{\mathbb{R}^n} g\,d\mu_i^i\right| + \left|\int_{\mathbb{R}^n} (g-f)\,d\mu_i^i\right| \\ &\leqslant M\|f-g\|_\infty + \varepsilon + M\|f-g\|_\infty \leqslant 2\varepsilon.\end{aligned}$$

This proves the claim. ■

(6) Finally, we remove the assumption $\sup_i \mu_i(\mathbb{R}^n) < \infty$. The assumption $\sup_i \mu_i(K) < \infty$ for every compact set $K \subset \mathbb{R}^n$ and the argument above show that for every j there is a subsequence (μ_i^j) of (μ_i^{j-1}) such that

$$\mu_i^j \lfloor_{B(0,j)} \rightharpoonup \nu^j, \quad j = 1, 2, \dots,$$

where ν^j is a Radon measure with $\nu^j(\mathbb{R}^n \setminus \overline{B}(0,j)) = 0$. The the diagonal sequence (μ_i^i) satisfies

$$\mu_i^i \lfloor_{B(0,j)} \rightharpoonup \nu^j \quad \text{as} \quad i \to \infty \quad \text{for every} \quad j = 1, 2, \dots.$$

Observe that $\nu^j \lfloor_{B(0,k)} = \nu^k$, $k = 1, \dots, j$. Thus we may define a Radon measure

$$\mu(A) = \sum_{j=2}^{\infty} \nu^j(A \cap (B(0,j) \setminus B(0,j+1))).$$

When j is so large that $\operatorname{supp} f \subset B(0,j)$, then

$$\int_{\mathbb{R}^n} f\,d\mu = \int_{\mathbb{R}^n} f\,d\nu^j = \lim_{i\to\infty} \int_{\mathbb{R}^n} f\,d\mu_i^i,$$

so that $\mu_i^i \rightharpoonup \mu$ as $i \to \infty$. □

5.5 Weak convergence in L^p.

Next we consider weak convergence in L^p. Recall that the Riesz representation theorem (Theorem 5.2) gives a characterization for $L^p(A)^*$ with $1 \leqslant p < \infty$.

Definition 5.18. Let $1 \leqslant p \leqslant \infty$. A sequence (f_i) of functions in $L^p(\mathbb{R}^n)$ converges weakly (weakly star if $p = \infty$) in $L^p(\mathbb{R}^n)$ to a function $f \in L^p(\mathbb{R}^n)$, if

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} f_i g\,dx = \int_{\mathbb{R}^n} f g\,dx \quad \text{for every} \quad g \in L^{p'}(\mathbb{R}^n).$$

Here we use the interpretation that $p' = \infty$ if $p = 1$ and $p' = 1$ if $p = \infty$.

Remark 5.19. $f_i \to f$ strongly in $L^p(\mathbb{R}^n)$ implies $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$.

Reason. By Hölder's inequality, we have

$$\begin{aligned} \left| \int_{\mathbb{R}^n} f_i g\,dx - \int_{\mathbb{R}^n} f g\,dx \right| &\leqslant \int_{\mathbb{R}^n} |f_i - f||g|\,dx \\ &\leqslant \|f_i - f\|_p \|g\|_{p'} \to 0 \quad \text{as} \quad i \to \infty \end{aligned}$$

for every $g \in L^{p'}(\mathbb{R}^n)$. ■

We illustrate some typical features of the behaviour of a sequence which converges weakly but not strongly.

Examples 5.20:

(1) (Oscillation) Let $f_i : (0,2\pi) \to \mathbb{R}$, $f_i(x) = \sin(ix)$, $i = 1,2,\dots$. Then f_i converges weakly to $f = 0$ in $L^p((0,2\pi))$, but $\|f_i\|_p = c(p) > 0$ for every $i = 1,2,\dots$, so that f_i does not converge to 0 in $L^p((0,2\pi))$. Observe, that

$$\|f\|_p < \liminf_{i\to\infty} \|f_i\|_p.$$

T H E M O R A L : Sequences of rapidly oscillating functions provide examples of weakly converging sequences that do not converge strongly.

(2) (Concentration) Let $f_i : (-1,1) \to \mathbb{R}$,

$$f_i(x) = \begin{cases} i, & 0 \leqslant x \leqslant 1/i, \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_i \to \delta_0$ weakly as measures in $(-1,1)$. Observe that f_i converges weakly to zero as measures in $(0,1)$, but f_i does not converge weakly in $L^1((0,1))$.

T H E M O R A L : Sequences of concentrating functions provide examples of weakly converging sequences that do not converge strongly.

(3) Let $1 < p < \infty$ and $f_i : (-1,1) \to \mathbb{R}$,

$$f_i(x) = \begin{cases} i^{1/p}, & 0 \leqslant x \leqslant 1/i, \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_i \to 0$ weakly in $L^p((-1,1))$, but the sequence (f_i) does not converge in $L^p((-1,1))$, since $\|f_i\|_p = 1$ for every $i = 1,2,\dots$. This shows that the norms $\|f_i\|_p = 1$ concentrate. However, $f_i \to 0$ in $L^q((-1,1))$ for every $q < p$. Indeed,

$$\int_{(-1,1)} |f_i|^q\,dx = i^{q/p-1} \to 0 \quad \text{as} \quad i \to \infty.$$

In particular, the norms $\|f_i\|_q$, $q < p$, do not concentrate.

The next result shows that any weakly converging sequence is bounded.

Theorem 5.21. If $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$, then

$$\|f\|_p \leqslant \liminf_{i\to\infty} \|f_i\|_p.$$

THE MORAL: The L^p-norm is lower semicontinuous with respect to the weak convergence.

Proof. $\boxed{1<p<\infty}$ If $\|f\|_p = 0$, the claim is clear. Hence we may assume that $\|f\|_p > 0$. The function $g = |f|^{p/p'} \operatorname{sign} f$ belongs $L^{p'}(\mathbb{R}^n)$, since

$$\|g\|_{p'} = \left(\int_{\mathbb{R}^n} |g|^{p'}\,dx\right)^{1/p'} = \left(\int_{\mathbb{R}^n} |f|^p\,dx\right)^{1/p'} = \|f\|_p^{p/p'} < \infty.$$

By the definition of the weak convergence

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} f_i g\,dx = \int_{\mathbb{R}^n} fg\,dx = \int_{\mathbb{R}^n} |f|^{p/p'} \underbrace{f \operatorname{sign} f}_{=|f|}\,dx = \int_{\mathbb{R}^n} |f|^p\,dx = \|f\|_p^p.$$

On the other hand, by Hölder's inequality

$$\left|\int_{\mathbb{R}^n} f_i g\,dx\right| \leqslant \|f_i\|_p \|g\|_{p'} = \|f_i\|_p \|f\|_p^{p/p'}.$$

Thus

$$\|f\|_p^p \leqslant \liminf_{i\to\infty} \|f_i\|_p \|f\|_p^{p/p'}.$$

Diving through by $\|f\|_p^{p/p'} > 0$ implies the claim.

$\boxed{p=1}$ Exercise.

$\boxed{p=\infty}$ Exhaust $\mathbb{R}^n$ with sets $A_j \subset A_{j+1}$, $A = \bigcup_{j=1}^\infty A_j$ with $|A_j| < \infty$ for every $j = 1,2,\ldots$. Let $\varepsilon > 0$ and set

$$A_{j,\varepsilon} = \{x \in A_j : |f(x)| \geqslant \|f\|_{L^\infty(A_j)} - \varepsilon\} \quad \text{and} \quad g = \chi_{A_{j,\varepsilon}} \operatorname{sign} f.$$

Then

$$\lim_{i\to\infty} \int_{A_j} f_i g\,dx = \int_{A_j} fg\,dx \geqslant (\|f\|_{L^\infty(A_j)} - \varepsilon)|A_{j,\varepsilon}|.$$

On the other hand, by Hölder's inequality

$$\left|\int_{A_j} f_i g\,dx\right| \leqslant \|f_i\|_\infty |A_{j,\varepsilon}|.$$

Diving through by $|A_{j,\varepsilon}| > 0$ implies

$$\|f\|_{L^\infty(A_j)} - \varepsilon \leqslant \liminf_{i\to\infty} \|f_i\|_{L^\infty(A_j)}.$$

The claim follows for the set A_j by letting $\varepsilon \to 0$.

The argument above gives

$$\begin{aligned}\|f\|_{L^\infty(A_j)} &\leqslant \liminf_{i\to\infty} \|f_i\|_{L^\infty(A_j)} \\ &\leqslant \liminf_{i\to\infty} \|f_i\|_{L^\infty(A)} \quad \text{for every} \quad j = 1,2,\ldots.\end{aligned}$$

The claim follows from this. □

Remark 5.22. Let $1 < p < \infty$. If $f_i \to f$ weakly in $L^p(A)$, it does not follow that that $\lim_{i\to\infty} \|f_i\|_p = \|f\|_p$. Nor does the reverse implication hold true. Example 5.20 (1) gives a counterexample for both claims. The following result explains the situation: If $f_i \to f$ weakly in $L^p(A)$ with $1 < p < \infty$ and $\lim_{i\to\infty} \|f_i\|_p = \|f\|_p$, then $f_i \to f$ strongly in $L^p(A)$. This result will not be proved here.

Remark 5.23. The previous theorem is a general fact in the functional analysis. Let X be a Banach space. If a sequence (x_i) converges weakly to $x \in X$, then it is bounded and

$$\|x\| \leqslant \liminf_{i\to\infty} \|x_i\|.$$

The previous theorem asserts that a weakly converging sequence is bounded. The next result shows that the converse is true up to a subsequence. One of the most useful applications of the weak convergence is in compactness arguments. A bounded sequence in L^p does not need to have any convergent subsequence with convergence interpreted in the standard L^p sense. However, there exists a weakly converging subsequence.

Theorem 5.24. Let $1 < p < \infty$. Assume that the sequence (f_i) of functions $f_i \in L^p(\mathbb{R}^n)$, $i = 1,2,\ldots$, satisfies

$$\sup_i \|f_i\|_p < \infty.$$

Then there exists a subsequence (f_{i_j}) and a function $f \in L^p(\mathbb{R}^n)$ such that $f_{i_j} \to f$ weakly in $L^p(\mathbb{R}^n)$.

THE MORAL: This shows that L^p with $1 < p < \infty$ is weakly sequentially compact, that is, every bounded sequence in L^p with $1 < p < \infty$ has a weakly converging subsequence. This is an analogue of the Bolzano-Weierstrass theorem.

Remark 5.25. The claim does not hold for $p = 1$. Indeed, if (f_i) is a sequence of nonnegative functions in $L^1(\mathbb{R}^n)$ with $\sup_i \|f_i\|_1 < \infty$, there is no guarantee that some subsequence will converge weakly in $L^1(\mathbb{R}^n)$. However, if μ is a Radon measure on $\mathbb{R}^n$ and consider the measures defined by

$$\nu_i(A) = \int_A f_i \, d\mu$$

for every Borel set $A \subset \mathbb{R}^n$. Then by Theorem 5.17, there exists a Radon measure ν on $\mathbb{R}^n$ such that

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} \varphi f_i \, d\mu = \int_{\mathbb{R}^n} \varphi \, d\mu \quad \text{for every} \quad \varphi \in C_0(\mathbb{R}^n).$$

Example 5.26. Let μ be the one-dimensional Lebesgue measure on $\mathbb{R}$. Then the sequence $f_i = i\chi_{[0,\frac{1}{i}]}$, $i = 1,2,\ldots$, converges in measure to zero, satisfies $\|f_i\|_1 = 1$

for every $i = 1, 2, \dots$, but does not converge weakly in $L^1(\mathbb{R})$. However, consider the measures

$$\nu_i(A) = i \int_{A \cap [0, \frac{1}{i}]} 1 \, dx = i \left| A \cap \left[0, \frac{1}{i}\right] \right|$$

for every Borel set $A \subset \mathbb{R}$, then $\nu_i \rightharpoonup \delta_0$ as $i \to \infty$.

Proof. (1) We may assume that $f_i \geqslant 0$ almost everywhere for every $i = 1, 2, \dots$, for otherwise we may consider the positive and negative parts f_i^+ and f_i^-. (Exercise)

(2) Define Radon measures μ_i, $i = 1, 2, \dots$, by setting

$$\mu_i(A) = \int_A f_i \, dx \tag{5.5}$$

where $A \subset \mathbb{R}^n$ is a Borel set. Then for each compact set $K \subset \mathbb{R}^n$,

$$\mu_i(K) = \int_K f_i \, dx \leqslant \left(\int_K f_i^p \, dx \right)^{1/p} |K|^{1-1/p} \quad \text{for every} \quad i = 1, 2, \dots.$$

This implies $\sup_i \mu_i(K) < \infty$. Thus we may apply Theorem 5.17 to find a Radon measure μ on $\mathbb{R}^n$ and a subsequence (μ_{i_j}), $j = 1, 2, \dots$, such that $\mu_i \rightharpoonup \mu$ as $i \to \infty$.

(3) μ is absolutely continuous with respect to the Lebesgue measure.

Reason. Assume that $A \subset \mathbb{R}^n$ is a bounded set with $|A| = 0$. Let $\varepsilon > 0$ and chose an open and bounded set $U \supset A$ such that $|U| < \infty$. Then

$$\begin{aligned}
\mu(U) &\leqslant \liminf_{i \to \infty} \mu_{i_j}(U) && \text{(Lemma 5.16)} \\
&= \liminf_{j \to \infty} \int_U f_{i_j} \, dx && (5.5) \\
&\leqslant \liminf_{j \to \infty} \left(\int_U f_{i_j}^p \, dx \right)^{1/p} |U|^{1-1/p} \leqslant c\varepsilon^{1-1/p}.
\end{aligned}$$

Thus $\mu(A) = 0$. ■

(4) By the Radon-Nikodym theorem 4.18, there exists $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ satisfying

$$\mu(A) = \int_A f \, dx \tag{5.6}$$

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

(4) $f \in L^p(\mathbb{R}^n)$.

Reason. Let $\varphi \in C_0(\mathbb{R}^n)$. Then

$$\begin{aligned}
\int_{\mathbb{R}^n} \varphi f \, dx &= \int_{\mathbb{R}^n} \varphi \, d\mu && (5.6) \\
&= \lim_{j \to \infty} \int_{\mathbb{R}^n} \varphi \, d\mu_{i_j} && (5.5) \\
&= \lim_{j \to \infty} \int_{\mathbb{R}^n} \varphi f_{i_j} \, dx && (5.6) \\
&\leqslant \sup_i \|f_I\|_p \|\varphi\|_q && \text{(Hölder)} \\
&\leqslant c\|\varphi\|_q. && \text{(assumption)}
\end{aligned}$$

By Theorem 5.2

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}^n} \varphi f\,dx : \varphi \in C_0(\mathbb{R}^n),\ \|\varphi\|_{p'} \leqslant 1\right\} < \infty.$$

■

(5) $f_{i_j} \to f$ weakly in $L^p(\mathbb{R}^n)$.

Reason. We showed above that

$$\lim_{j\to\infty}\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx = \int_{\mathbb{R}^n} \varphi f\,dx$$

for every $\varphi \in C_0(\mathbb{R}^n)$. Assume that $g \in L^{p'}(\mathbb{R}^n)$. Let $\varepsilon > 0$ and choose $\varphi \in C_0(\mathbb{R}^n)$ with $\|g-\varphi\|_{p'} < \varepsilon$. Then

$$\begin{aligned}
&\left|\int_{\mathbb{R}^n} f_{i_j} g\,dx - \int_{\mathbb{R}^n} f g\,dx\right| \\
&\leqslant \left|\int_{\mathbb{R}^n} f_{i_j} g\,dx - \int_{\mathbb{R}^n} \varphi f_{i_j}\,dx\right| + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| + \left|\int_{\mathbb{R}^n} \varphi f\,dx - \int_{\mathbb{R}^n} g f\,dx\right| \\
&\leqslant \|f_{i_j}\|_p \|g-\varphi\|_{p'} + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| + \|f\|_p \|g-\varphi\|_{p'} \\
&\leqslant \varepsilon \sup_i \|f_i\|_p + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| + \varepsilon\|f\|_p
\end{aligned}$$

This implies

$$\lim_{j\to\infty}\int_{\mathbb{R}^n} f_{i_j} g\,dx = \int_{\mathbb{R}^n} \varphi\,dx.$$

■

Remark 5.27. There is a general theorem in functional analysis, which says that a Banach space is weakly sequentially compact if and only if it is reflexive. This is another manifestation that $L^p(\mathbb{R}^n)$ spaces are reflexive for $1 < p < \infty$.

Corollary 5.28. Let $1 < p \leqslant \infty$. Assume that the sequence (f_i) of functions $f_i \in L^p(\mathbb{R}^n)$, $i = 1, 2, \ldots$. Then $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ if and only if

(1) $\sup_i \|f_i\|_p < \infty$ and
(2) $\lim_{i\to\infty}\int_A f_i\,dx = \int_A f\,dx$ for every measurable set $A \subset \mathbb{R}^n$ with $\mu(A) < \infty$.

Proof. The result follows from the fact that simple functions are dense in $L^{p'}(\mathbb{R}^n)$ using the results above. □

T H E M O R A L : Property (2) asserts that the averages of the functions f_i converge to the average of f over A.

THE END