JUHA KINNUNEN

Real Analysis

Department of Mathematics, Aalto University
2024

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 that L^p is a complete space 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 quantitative information on the average size of measurable functions and boundedness 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 all play an important role in in Fourier analysis, 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 $f : A \to [-\infty, \infty]$ a μ-measurable function. Then $f \in L^p(A)$, $1 \leqslant p < \infty$, if

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

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 $A = B(0,1) = \{x \in \mathbb{R}^n : |x| < 1\}$, $\Omega_n = |B(0,1)|$ and denote $A_i = B(0,2^{-i+1}) \setminus B(0,2^{-i})$, $i = 1,2,\dots$. Here $|A|$ denotes the n-dimensional Lebesgue outer measure of $A \subset \mathbb{R}^n$. Then

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

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

On the other hand,

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

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

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

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

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

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

$$\int_{\mathbb{R}^n} f(|x|)\,dx = \omega_{n-1}\int_0^\infty f(r)r^{n-1}\,dr, \tag{1.5}$$

where

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

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 $B(x,r) = \{y\in\mathbb{R}^n : |y-x|<r\}$, with $x\in\mathbb{R}^n$ and $r>0$. By the translation and scaling invariance, we have

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

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

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

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,

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

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

On the other hand, the integrals can be computer directly by (1.5). 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

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

Remarks 1.6:
Formula (1.5) 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 1.7. If $f\in L^p(A)$, then $|f(x)|<\infty$ for μ-almost every $x\in A$.

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

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

By Chebyshev's inequality

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

∎

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

Remark 1.8. If $f\in L^p(A)$, then $\{x\in\mathbb{R}^n: |f(x)|\neq 0\}$ is σ-finite with respect to μ.

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

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

By Chebyshev's inequality

$$\mu(A_i) = \mu(\{x\in A: |f(x)|\geqslant \tfrac{1}{i}\}) = \int_{A_i} 1\,d\mu$$
$$\leqslant i^p\int_{A_i}|f|^p\,d\mu < \infty \qquad (|f|\geqslant \tfrac{1}{i} \text{ in } A_i)$$

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

∎

1.2 L^p norm

Let $A \subset \mathbb{R}^n$ a μ-measurable set and $1 \leqslant p < \infty$. The L^p norm of $f \in L^p(A)$ is the number

$$\|f\|_p = \|f\|_{L^p(A)} = \left(\int_A |f|^p \, d\mu\right)^{\frac{1}{p}}.$$

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

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

(2) $\|f\|_p = 0 \iff f = 0$ 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.

$\boxed{\Longleftarrow}$ 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 $\|f\|_p = 0$.

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

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

By Chebyshev's inequality

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

Thus $\mu(A_i) = 0$ for every $i = 1,2,\dots$ 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.

For μ-measurable functions f and g on a μ-measurable set A, we are interested in the condition $f(x) = g(x)$ for μ-almost every $x \in A$, which 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 formal and consider the equivalence relation

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

but this is not 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. 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 1.9.

(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. $\boxed{(1)}$ $\displaystyle\int_A |af|^p \, d\mu = |a|^p \int_A |f|^p \, d\mu < \infty.$

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

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

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

$$\begin{aligned}(a + b)^p &\leqslant (2 \max\{a, b\})^p = 2^p \max\{a^p, b^p\} \\ &\leqslant 2^p (a^p + b^p), \quad a, b \geqslant 0, \quad 0 < p < \infty \end{aligned} \tag{1.10}$$

implies that

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

□

Remark 1.11. 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 with the L^p norm only for $p \geqslant 1$ as we shall see later.

Remark 1.12. A more careful analysis gives the useful inequality

$$(a+b)^p \leqslant 2^{p-1}(a^p+b^p), \quad a,b \geqslant 0, \quad 1 \leqslant p < \infty. \tag{1.13}$$

Remarks 1.14:

(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

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

(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

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

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

Definition 1.15. 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 1.16. 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 1.17 (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 $a^p = b^{p'}$.

THE MORAL: 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

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

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

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

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 $\varphi'(t) = 0 \Longleftrightarrow t = 1$, 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 1.18:

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

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

(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 1.19 (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

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

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 $x \in A$.

THE MORAL: Hölder's inequality is very useful tool in estimating a product of functions.

Remark 1.20. 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 $\|g\|_{p'} = 0$. The result is also clear if $\mu(A) = 0$. Thus we may assume that $\|f\|_p > 0$, $\|g\|_{p'} > 0$ and $\mu(A) > 0$. Let

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

Then

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

By Young's inequality

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

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

$$\frac{1}{p}|\widetilde{f}|^p + \frac{1}{p'}|\widetilde{g}|^{p'} - |\widetilde{f}\widetilde{g}| = 0$$

μ-almost everywhere in A. An equality occurs in Young's inequality if and only if $|\widetilde{f}|^p = |\widetilde{g}|^{p'}$ μ-almost everywhere in A. In this case, we have

$$|f(x)|^p = \frac{\|f\|_p^p}{\|g\|_{p'}^{p'}}|g(x)|^{p'}$$

for μ-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 1.21:

(1) For $p = 2$ we have the Cauchy-Schwarz inequality

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

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

Lemma 1.22 (Jensen's inequality). Let $1 \leqslant p < q < \infty$ and assume that $A \subset \mathbb{R}^n$ is a μ-measurable set with $0 < \mu(A) < \infty$. Then

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

THE MORAL: An integral average is an increasing function of the power.

Proof. By Hölder's inequality with the exponents $\frac{q}{p}$ and $\left(\frac{q}{p}\right)' = \frac{q}{q-p}$, we have

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

□

Remark 1.23. If $1 \leqslant p < q < \infty$ and $\mu(A) < \infty$, then $L^q(A) \subset L^p(A)$.

WARNING: 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

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

Assume that $1\leqslant p<q<\infty$. Choose b such that $\frac{1}{q}\leqslant b<\frac{1}{p}$. 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 1.24:

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

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

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 $L^q((0,1))$ 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 exist 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 $a_i\geqslant 0$ with $a_i\to\infty$ as $i\to\infty$ 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 exist 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 $a_i\geqslant 0$ with $a_i\to 0$ as $i\to\infty$ 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 $f\in L^q(A)\setminus L^p(A)$. 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 1.25. There is a more general version of Jensen's inequality. Assume that $A \subset \mathbb{R}^n$ is a μ-measurable set with $0 < \mu(A) < \infty$. 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 1.26 (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 1.27. Elementary inequalities (1.13) and (1.30) imply that

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

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 1.9, 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 $\|f+g\|_p > 0$. By Hölder's inequality, we have

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

Since $(p-1)p' = p$ and $0 < \|f+g\|_p < \infty$, we have

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

It remains to consider when an equality occurs. This happens if there is an equality in the pointwise inequality

$$|f(x)+g(x)|^p = |f(x)+g(x)|^{p-1}|f(x)+g(x)| \leqslant |f(x)+g(x)|^{p-1}(|f(x)|+|g(x)|)$$

for μ-almost every $x \in A$ as well as an equality in the application of Hölder's inequality. An equality occurs in Hölder's inequality if

$$c_1|f(x)|^p = |f(x)+g(x)|^p = c_2|g(x)|^p$$

for μ-almost every $x \in A$. Equality occurs in in the pointwise inequality above if $f(x)$ and $g(x)$ have the same sign. This completes the proof. □

Remark 1.28. It is possible to prove Minkowski's inequality directly by Young's inequality instead of Hölder's inequality (exercise).

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 1.9. However, $\|f\|_p$ is not a norm for $0 < p < 1$.

Reason. Let $f,g:\mathbb{R}\to\mathbb{R}$, $f=\chi_{[0,\frac{1}{2})}$ and $g=\chi_{[\frac{1}{2},1]}$. Then $f+g=\chi_{[0,1]}$ so that $\|f+g\|_p=1$. On the other hand,

$$\|f\|_p=2^{-\frac{1}{p}} \quad \text{and} \quad \|g\|_p=2^{-\frac{1}{p}}.$$

Thus

$$\|f\|_p+\|g\|_p=2\cdot 2^{-\frac{1}{p}}=2^{1-\frac{1}{p}}<1,$$

when $0<p<1$. 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 1.29. 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

$$(a+b)^p\leqslant a^p+b^p,\quad a,b\geqslant 0,\quad 0<p<1, \tag{1.30}$$

implies that

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

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 satisfy the triangle inequality.

Remarks 1.31:

(1) By (1.10), we have

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

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 1.32. 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

$$0<\int_A|f|^p\,d\mu<\infty.$$

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

1.4 Completeness of L^p

Next we prove a famous theorem, which is not only important in the theory of L^p spaces, but has a 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 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

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

W A R N I N G : 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).

C L A I M : If $f_i \to f$ in $L^p(A)$ as $i \to \infty$, then (f_i) is a Cauchy sequence in $L^p(A)$.

Reason. Let $\varepsilon > 0$. Since $f_i \to f$ in $L^p(A)$ as $i \to \infty$, there exists i_ε such that $\|f_i - f\|_p < \frac{\varepsilon}{2}$ when $i \geqslant i_\varepsilon$. Minkowski's inequality implies that

$$\|f_i - f_j\|_p \leqslant \|f_i - f\|_p + \|f - f_j\|_p < \tfrac{\varepsilon}{2} + \tfrac{\varepsilon}{2} = \varepsilon$$

when $i,j \geqslant i_\varepsilon$. ∎

Theorem 1.33 (Riesz-Fischer). For every Cauchy sequence (f_i) in $L^p(A)$, $1 \leqslant p < \infty$, there exists $f \in L^p(A)$ such that $f_i \to f$ in $L^p(A)$ as $i \to \infty$.

T H E M O R A L : $L^p(A)$, $1 \leqslant p < \infty$, 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 construct a subsequence as follows. Choose i_1 such that

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

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

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

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

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

Let

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

Then

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

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

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

Thus $g \in L^p(A)$ and consequently $g(x) < \infty$ for μ-almost every $x \in A$. It follows that the telescoping 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 (f_{i_k}) which converges μ-almost everywhere in A. Next we show that the original sequence converges to f in $L^p(A)$.

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

Reason. Let $\varepsilon > 0$ and let (f_{i_k}) be a subsequence which converges to f μ-almost everywhere in A. Since (f_i) is a Cauchy sequence in $L^p(A)$, there exists i_ε such that $\|f_{i_k} - f_i\|_p < \varepsilon$ when $i, i_k \geqslant i_\varepsilon$. For a fixed $i \geqslant i_\varepsilon$, 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

$$\begin{aligned}\left(\int_A |f - f_i|^p \, d\mu\right)^{\frac{1}{p}} &= \left(\int_A \liminf_{k\to\infty} |f_{i_k} - f_i|^p \, d\mu\right)^{\frac{1}{p}} \\ &\leqslant \liminf_{k\to\infty} \left(\int_A |f_{i_k} - f_i|^p \, d\mu\right)^{\frac{1}{p}} \leqslant \varepsilon. \end{aligned}$$ ■

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 $\|f_i - f\|_p \leqslant \varepsilon$ when $i \geqslant i_\varepsilon$. 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 part of the proof of the Riesz-Fisher theorem, which we now state.

Corollary 1.34. 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. □

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

Remarks 1.35:

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

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

Reason. By Minkowski's inequality

$$\|f_i\|_p = \|f_i - f + f\|_p \leqslant \|f_i - f\|_p + \|f\|_p,$$

which implies $\|f_i\|_p - \|f\|_p \leqslant \|f_i - f\|_p$. By switching the roles of f and f_i, we have $\|f\|_p - \|f_i\|_p \leqslant \|f_i - f\|_p$. Thus

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

from which it follows that

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

■

(2) If $f_i \to f$ in $L^p(A)$ as $i \to \infty$, then $f_i \to f$ in measure.

Reason. By Chebyshev's inequality

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

■

(3) If $f_i \to f$ in measure as $i \to \infty$, 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 $L^1(A)$ as $i \to \infty$ 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.

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

■

The following example shows that pointwise convergence almost everywhere does not imply L^p convergence and L^p convergence does not imply pointwise convergence almost everywhere.

Example 1.36. In the following examples we assume that μ is the Lebesgue measure.

(1) $f_i \to f$ almost everywhere as $i \to \infty$ does not imply $f_i \to f$ in L^p. Let $f_i : \mathbb{R} \to \mathbb{R}$, $f_i(x) = \chi_{[i-1,i)}(x)$, $i = 1, 2, \ldots$, and $f = 0$. Assume that μ is the Lebesgue measure. Then $f_i(x) \to 0$ for every $x \in \mathbb{R}$. 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 $L^p(\mathbb{R})$, $1 \leqslant p < \infty$.

(2) $f_i \to f$ almost everywhere as $i \to \infty$ does not imply $f_i \to f$ in L^p. Let $f_i : \mathbb{R} \to \mathbb{R}$,

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

Then

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

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

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

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

(3) $f_i \to f$ in L^p as $i \to \infty$ does not imply $f_i \to f$ almost everywhere. Consider the sliding sequence of functions $f_i : \mathbb{R} \to \mathbb{R}$,

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

Then

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

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 $k \to \infty$.

(4) A sequence can converge in L^p without converging in L^q. Consider $f_i : \mathbb{R} \to \mathbb{R}$,

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

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

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

Theorem 1.37. Let $1 \leqslant p < \infty$. Assume that $f_i \in L^p(A)$, $i = 1,2,\dots$, $f_i \to f$ μ-almost everywhere in A as $i \to \infty$. If there exists $g \in L^p(A)$, $g \geqslant 0$, such that $|f_i| \leqslant g$ μ-almost everywhere in A for every $i = 1,2,\dots$, then $f \in L^p(A)$ and $f_i \to f$ in $L^p(A)$ as $i \to \infty$.

THE MORAL: Pointwise convergence implies the norm convergence in L^p if the sequence is uniformly bounded by a function in L^p. This is a dominated convergence theorem in L^p.

Proof. Since $f_i \to f$ μ-almost everywhere in A as $i \to \infty$ and $|f_i| \leqslant g$ μ-almost everywhere in A for every $i = 1, 2, \dots$, we conclude that

$$|f| = |\lim_{i\to\infty} f_i| = \lim_{i\to\infty} |f_i| \leqslant g$$

μ-almost everywhere in A. Moreover, we have

$$|f_i - f|^p \leqslant (|f_i| + |f|)^p \leqslant (g+g)^p = 2^p g^p \in L^1(A)$$

μ-almost everywhere in A. By the dominated convergence theorem

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

□

Theorem 1.38. 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 (1.10), we have

$$2^p(|f_i|^p + |f|^p) - |f_i - f|^p \geqslant 0$$

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

$$\lim_{i\to\infty} (2^p(|f_i|^p + |f|^p) - |f_i - f|^p) = 2^{p+1}|f|^p$$

μ-almost everywhere in A. By Fatou's lemma, we obtain

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

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

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

Thus

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

□

Remark 1.39. Let $1 \leqslant p < \infty$. Assume that $f_i \in L^p(A)$, $0 \leqslant f_i \leqslant f_{i+1}$ μ-almost everywhere in A, $i = 1,2,\dots$. Then the pointwise limit $f = \lim_{i\to\infty} f_i$ exists μ-almost everywhere in A. The monotone convergence theorem implies that

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

Theorem 1.38 (or Theorem 1.37) implies $f_i \to f$ in $L^p(A)$ as $i \to \infty$.

THE MORAL: An increasing sequence of nonnegative functions in L^p converges in L^p, if the limit function belongs to L^p. This is a monotone convergence theorem in L^p.

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 functions on a set of measure zero.

Definition 1.40. 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. 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 $\|f\|_\infty < \infty$. Note also that if $f \in L^\infty(A)$, then there exists M, $0 \leqslant M < \infty$, such that $|f(x)| \leqslant M$ for μ-almost every $x \in A$. This implies

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

and thus $\operatorname{ess\,sup}_{x\in A} f(x) < \infty$ and $\operatorname{ess\,inf}_{x\in A} f(x) > -\infty$.

T H E M O R A L : L^∞ consists of measurable functions that can be redefined on a set of measure zero so that the functions become bounded. The essential supremum is supremum outside sets of measure zero. Observe that the standard supremum of a bounded function f is

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

W A R N I N G : 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 1.41. 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 $l^\infty = \big\{(x_i) : \sup_{i\in\mathbb{N}} |x_i| < \infty\big\}$. Here (x_i) is a sequence of real (or complex) numbers. Thus l^∞ is the space of bounded sequences.

Example 1.42. 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 $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \frac{1}{|x|}$. Then $f(x) < \infty$ for almost every $x \in \mathbb{R}^n$, but $f \notin L^\infty(\mathbb{R}^n)$.

Remarks 1.43:

(1) $\|f\|_\infty \leqslant \sup_{x\in A} |f(x)|$.
(2) Let $f \in L^\infty(A)$. By the definition of infimum, for every $\varepsilon > 0$, we have

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

Lemma 1.44. Assume that $f \in L^\infty(A)$. Then

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

T H E M O R A L : If $f \in L^\infty(A)$, then $|f(x)| \leqslant \|f\|_\infty$ for μ-almost every $x \in A$.

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

$$M_i < \operatorname*{ess\,sup}_{y\in A} f(y) + \tfrac{1}{i}$$

and $f(x) \leqslant M_i$ for μ-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

$$f(x) \leqslant M_i < \operatorname*{ess\,sup}_{y\in A} f(y) + \tfrac{1}{i} \quad \text{for every} \quad x \in A \setminus N, \quad i = 1, 2, \dots .$$

Letting $i \to \infty$, we obtain $f(x) \leqslant \operatorname{ess\,sup}_{y\in A} f(y)$ for every $x \in A \setminus N$.

(2) (Exercise) □

Remark 1.45. Let $f \in L^\infty(A)$. Lemma 1.44 implies that

$$\|f\|_\infty = \min\{M : |f(x)| \leqslant M \text{ for } \mu\text{-almost every } x \in A\}.$$

Note that $\inf B = \min B \iff \inf B \in B$ where $B \subset \mathbb{R}$ is bounded from below.

Lemma 1.46 (Minkowski's inequality for $p = \infty$). If $f, g \in L^\infty(A)$, then

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

T H E M O R A L : 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.

Proof. By Lemma 1.44, we have $|f(x)| \leqslant \|f\|_\infty$ for μ-almost every $x \in A$ and $|g(x)| \leqslant \|g\|_\infty$ for μ-almost every $x \in A$. Thus

$$|f(x) + g(x)| \leqslant |f(x)| + |g(x)| \leqslant \|f\|_\infty + \|g\|_\infty$$

for μ-almost every $x \in A$. By the definition of the L^∞ norm, we have $\|f + g\|_\infty \leqslant \|f\|_\infty + \|g\|_\infty$. □

Theorem 1.47 (Hölder's inequality for $p = \infty$ and $p' = 1$). If $f \in L^1(A)$ ja $g \in L^\infty(A)$, then $fg \in L^1(A)$

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

T H E M O R A L : In practice, we take the essential supremum out of the integral.

Proof. By Lemma 1.44, we have $|g(x)| \leqslant \|g\|_\infty$ for μ-almost every $x \in A$. This implies

$$|f(x)g(x)| \leqslant \|g\|_\infty |f(x)|$$

for μ-almost every $x \in A$ and thus

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

□

Remark 1.48. There is also an L^p version $\|fg\|_p \leqslant \|g\|_\infty \|f\|_p$ of the previous theorem.

Next result justifies the notation $\|f\|_\infty$ as a limit notion of $\|f\|_p$ as $p \to \infty$.

Theorem 1.49. If $f \in L^q(A)$ for some $1 \leqslant q < \infty$, 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 $p \to \infty$. This gives a method to pass from average the information $\|f\|_p$ to the pointwise information $\|f\|_\infty$ outside a set of measure zero.

Proof. Assume that $\|f\|_q < \infty$ for some $1 \leqslant q < \infty$ and let $p > q$. Let

$$A_\lambda = \{x \in A : |f(x)| > \lambda\}, \quad \lambda \geqslant 0.$$

Assume that $0 \leqslant \lambda < \|f\|_\infty$. 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 $\|f\|_p \geqslant \lambda \mu(A_\lambda)^{\frac{1}{p}}$. 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 $1 \leqslant q < p < \infty$, we have

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

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 1.50:

(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

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

is an increasing function of p.

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

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

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

and

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

Theorem 1.51. $L^\infty(A)$ is a Banach space.

THE MORAL: 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 1.44, we have

$$|f_i(x)-f_j(x)| \leqslant \|f_i - f_j\|_\infty$$

for μ-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 $k=1,2,\dots$, there exists i_k such that

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

This implies

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

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

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

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

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

which implies

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

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 1.52. 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 $\mu(N) = 0$. This means that L^∞ convergence is uniform convergence outside a set of measure zero. Uniform convergence outside a set of measure zero implies immediately pointwise convergence almost everywhere, compare to Corollary 1.34 for L^p with $1 \leqslant p < \infty$.

Example 1.53. 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 $i = 1, 2, \ldots$ 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 $i = 1, 2, \ldots$, $\|f\|_\infty = 1$ so that

$$\lim_{i \to \infty} \|f_i\|_\infty = \|f\|_\infty,$$

but $\|f_i - f\|_\infty = 1$ for every $i = 1, 2, \ldots$. Thus

$$\lim_{i \to \infty} \|f_i - f\|_\infty = 1 \neq 0.$$

This shows that the claim of Theorem 1.38 does not hold when $p = \infty$.

1.6 Density of continuous functions

We discuss approximation of L^p functions by compactly supported continuous functions. We assume that the underlying measure μ is the Lebesgue measure on $\mathbb{R}^n$ throughout this section.

Definition 1.54. 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 1.55. Let $f,g:\mathbb{R}^n \to \mathbb{R}$. Support of a function of has the following properties (exercise):

(1) $\operatorname{supp}(f+g) \subset \operatorname{supp} f \cup \operatorname{supp} g$,
(2) $\operatorname{supp}(af) = \operatorname{supp} f$, if $a \neq 0$ and
(3) $\operatorname{supp}(fg) \subset \operatorname{supp} f \cap \operatorname{supp} g$.

Remark 1.56. $C_0(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ 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.

$\boxed{p = \infty}$

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

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

Theorem 1.57. 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 $L^\infty(\mathbb{R}^n)$.

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

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

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

Observe that

$$|f_i - f|^p \leqslant (|f_i| + |f|)^p \leqslant 2^p |f|^p, \quad i = 1,2,\dots.$$

Since $|f|^p \in L^1(\mathbb{R}^n)$, by the dominated convergence theorem (Theorem 1.37), we have

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

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.

(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.

(3) Since $f\geqslant 0$ is measurable, there exists an increasing sequence of simple functions s_i such that

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

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

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

and thus $s_i\in L^p(\mathbb{R}^n)$, $i=1,2,\dots$. Observe that

$$|s_i-f|^p\leqslant(|s_i|+|f|)^p\leqslant 2^p|f|^p,\quad i=1,2,\dots.$$

Since $|f|^p\in L^1(\mathbb{R}^n)$, by the dominated convergence theorem (Theorem 1.37), we have

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

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 an approximation result for measurable sets, there exist an open set $G\supset A$ and a closed set $F\subset A$ such that $|G\setminus F|^{\frac{1}{p}}<\varepsilon$, 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

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

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

$$g(x) = \frac{\operatorname{dist}(x, \mathbb{R}^n \setminus U)}{\operatorname{dist}(x, F) + \operatorname{dist}(x, \mathbb{R}^n \setminus U)}, \tag{1.58}$$

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

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

Remarks 1.59:

(1) The proof above shows that $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n; \mu)$, $1 \leqslant p < \infty$, for every Radon measure μ on $\mathbb{R}^n$.

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

Remark 1.60. 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. The spaces $L^p(\mathbb{R}^n)$ with $1 \leqslant p < \infty$ are separable, since the collection of rational linear combinations of the characteristic functions of those sets that are finite unions of intervals with rational endpoints (or dyadic cubes) gives a countable dense subset (exercise). However, for other measures than the Lebesgue measure separability depends on the measure. The space $L^\infty(\mathbb{R}^n)$ is not separable, since $\{\chi_{B(x,r)} : x \in \mathbb{R}^n, r > 0\} \subset L^\infty(\mathbb{R}^n)$ is an uncountable set, but there does not exist a countable dense subset. Observe that $\|\chi_{B(x,r)} - \chi_{B(x,s)}\|_\infty = 1$ for $r \neq s$ (exercise).

1.7 Continuity of translation

We assume that the underlying measure μ is the Lebesgue measure on $\mathbb{R}^n$ throughout this section. We discuss a useful continuity property of the integral. This result 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). For the Fourier transform, see [7, Chapter 13].

Theorem 1.61. 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

$$\lim_{y\to 0} \|\tau_y f - f\|_p = 0.$$

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 1.57, there exists $g \in C_0(\mathbb{R}^n)$ such that

$$\left(\int_{\mathbb{R}^n} |f(x) - g(x)|^p \, dx\right)^{\frac{1}{p}} < \frac{\varepsilon}{3}.$$

The translation invariance of the Lebesgue integral implies that

$$\left(\int_{\mathbb{R}^n} |f(x+y) - g(x+y)|^p \, dx\right)^{\frac{1}{p}} = \left(\int_{\mathbb{R}^n} |f(x) - g(x)|^p \, dx\right)^{\frac{1}{p}}.$$

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. Thus there exists $0 < \delta \leqslant 1$ such that

$$|g(x+y) - g(x)| < \frac{\varepsilon}{3|B(0,r+1)|^{\frac{1}{p}}} \quad \text{for every} \quad x \in \mathbb{R}^n, \quad |y| < \delta.$$

Since $g(x+y)-g(x)=0$ for every $x \in \mathbb{R}^n \setminus B(0,r+1)$, we have

$$\begin{aligned}\left(\int_{\mathbb{R}^n} |g(x+y)-g(x)|^p\,dx\right)^{\frac{1}{p}} &= \left(\int_{B(0,r+1)} |g(x+y)-g(x)|^p\,dx\right)^{\frac{1}{p}} \\ &< \frac{\varepsilon}{3|B(0,r+1)|^{\frac{1}{p}}}|B(0,r+1)|^{\frac{1}{p}} = \frac{\varepsilon}{3}.\end{aligned}$$

Therefore

$$\begin{aligned}&\left(\int_{\mathbb{R}^n} |f(x+y)-f(x)|^p\,dx\right)^{\frac{1}{p}} \\ &\quad\leqslant \left(\int_{\mathbb{R}^n} |f(x+y)-g(x+y)|^p\,dx\right)^{\frac{1}{p}} + \left(\int_{\mathbb{R}^n} |g(x+y)-g(x)|^p\,dx\right)^{\frac{1}{p}} \\ &\qquad + \left(\int_{\mathbb{R}^n} |g(x)-f(x)|^p\,dx\right)^{\frac{1}{p}} \\ &\quad< \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\varepsilon.\end{aligned}$$

□

1.8 Local L^p space

If we are interested in pointwise properties of functions, it is not necessary to require integrablity conditions over the whole underlying domain. We assume that the underlying measure μ is the Lebesgue measure on $\mathbb{R}^n$ throughout this section.

Definition 1.62. Let $\Omega \subset \mathbb{R}^n$ be an open set and assume that $f : \Omega \to [-\infty,\infty]$ is a measurable function. Then $f \in L^p_{\mathrm{loc}}(\Omega)$, 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 1.63:
$L^p(\Omega) \subset L^p_{\mathrm{loc}}(\Omega)$, 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 $f \in L^p_{\mathrm{loc}}(\mathbb{R}^n)$ for every $1 \leqslant p < \infty$.

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

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

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

(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 1.64:

(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

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

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

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

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 $\mathbb{R}^n$.

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 Definition of the maximal function

We begin with the definition of the maximal function.

Definition 2.1. 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 $x \in \mathbb{R}^n$.

THE MORAL: The Hardy-Littlewood maximal function gives the maximal integral average of the absolute value of the function on balls centered at a point. As we shall see later, the maximal function is used to give bounds for other more complicated operators. Instead of the precise value at a given point, we are interested in estimates for the maximal function.

Remarks 2.2:

(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 $x \in \mathbb{R}^n$.

Reason. Let $x \neq 0$ and $r > 2|x|$. Then $B(0,\frac{r}{2}) \subset B(x,r)$. To see this, let $y \in B(0,\frac{r}{2})$. Then $|y-x| \leqslant |y| + |x| < \frac{r}{2} + \frac{r}{2} = r$. It follows that $y \in B(x,r)$. Thus we have

$$\begin{aligned}
\frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\,dy &\geqslant \frac{1}{|B(x,r)|} \int_{B(0,\frac{r}{2})} |y|\,dy \\
&= \frac{1}{\Omega_n r^n} \omega_{n-1} \int_0^{\frac{r}{2}} \rho \rho^{n-1}\,d\rho \\
&= \frac{n\Omega_n}{\Omega_n r^n} \int_0^{\frac{r}{2}} \rho^n\,d\rho \\
&= \frac{n}{r^n} \frac{1}{n+1} \left(\frac{r}{2}\right)^{n+1} \\
&= \frac{n}{n+1} \frac{r}{2^{n+1}} \xrightarrow{r\to\infty} \infty.
\end{aligned}$$

■

(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 2.3:

(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 $z = \frac{1}{2}(x+a)$ and $r = \frac{1}{2}(x-a)$. Similarly, when $x < a$, the maximal average is obtained when $z = \frac{1}{2}(b+x)$ and $r = \frac{1}{2}(b-x)$. 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 2.4. 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 larger 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 $|x - y| < \delta$. It follows that

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

if $0 < r \leqslant \delta$. Thus

$$|f(x)| = \lim_{r\to 0}\frac{1}{|B(x,r)|}\int_{B(x,r)}|f(y)|\,dy \leqslant Mf(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

□

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

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

Since the volume of a ball is a continuous function of the radius of a ball, we have

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

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. □

Example 2.6. Let $R > 0$ and $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \chi_{B(0,R)}(x)$ for every $x \in \mathbb{R}^n$. Then $Mf(x) = 1$ for every $x \in B(0,R)$ and $Mf(x) < 1$ for every $x \in \partial B(0,R)$. Thus Mf is not continuous on $\partial B(0,R)$.

Reason. Let $x \in B(0,R)$. Since $B(0,R)$ is open, there exists $r > 0$ such that $B(x,r) \subset B(0,R)$. Thus

$$Mf(x) \geqslant \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\,dy = 1.$$

On the other hand, we have

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

for every $r > 0$. This shows that

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

This proves the first claim.

To prove the second claim, let $x \in \partial B(0,R)$ and $r > 0$. Then there exists $y \in B(x,r) \setminus B(0,R)$ such that $B(y,\frac{r}{2}) \subset B(x,r) \setminus B(0,R)$. Thus

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

where

$$\begin{aligned} |B(x,r)\cap B(0,R)| &= |B(x,r)| - |B(x,r) \setminus B(0,R)| \\ &\leqslant |B(x,r))| - |B(y,\tfrac{r}{2})| = |B(x,r))| - |B(x,\tfrac{r}{2})| \\ &= |B(x,r))| - \left(\frac{1}{2}\right)^n |B(x,r)| = \frac{2^n-1}{2^n}|B(x,r)|. \end{aligned}$$ ■

This shows that

$$Mf(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\,dy \leqslant \frac{2^n-1}{2^n} < 1$$

for every $x \in \partial B(0,R)$.

2.2 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 2.7. 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 $\tau_y f(x) = f(x+y)$.

(5) (Scaling invariance) $M(\delta_a f)(x) = (\delta_a Mf)(x)$, where $\delta_a f(x) = f(ax)$ with $a > 0$.

Proof. Exercise. □

Example 2.8. Let $0 < \alpha < n$ and define $f : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$, $f(x) = |x|^{-\alpha}$. Let $x \in \mathbb{R}^n \setminus \{0\}$ and write $z = \frac{x}{|x|}$. By Lemma 2.7 (5) and (3), we have

$$\begin{aligned} Mf(x) &= Mf(|x|z) = (\delta_{|x|} Mf)(z) \\ &= M(\delta_{|x|} f)(z) = |x|^{-\alpha} Mf(z). \end{aligned}$$

Thus $Mf(x) = Mf(z)|x|^{-\alpha}$ for every $x \in \mathbb{R}^n \setminus \{0\}$, where $z \in \partial B(0,1)$. Since f is radial, the value $Mf(z)$ is independent of $z \in \partial B(0,1)$. Moreover,

$$\begin{aligned} Mf(z) &= \sup_{r>0} \frac{1}{|B(z,r)|} \int_{B(z,r)} |f(y)|\,dy \\ &\leqslant \sup_{0<r<\frac{1}{2}} \frac{1}{|B(z,r)|} \int_{B(z,r)} |f(y)|\,dy + \sup_{r \geqslant \frac{1}{2}} \frac{1}{|B(z,r)|} \int_{B(z,r)} |f(y)|\,dy \\ &\leqslant \left(\tfrac{1}{2}\right)^{-\alpha} + C(n) \sup_{r \geqslant \frac{1}{2}} \frac{1}{|B(0,r+1)|} \int_{B(0,r+1)} |y|^{-\alpha}\,dy \\ &\leqslant 2^{\alpha} + c(n) \sup_{r \geqslant \frac{1}{2}} \frac{1}{(r+1)^n} \int_0^{r+1} \rho^{-\alpha} \rho^{n-1}\,d\rho \\ &\leqslant 2^{\alpha} + c(n,\alpha) \sup_{r \geqslant \frac{1}{2}} (r+1)^{-\alpha} < \infty. \end{aligned}$$

This shows that Mf is a constant multiple of f.

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. We begin with an L^∞ estimate, which follows directly form the definitions.

Lemma 2.9. If $f \in L^\infty(\mathbb{R}^n)$, then $Mf \in L^\infty(\mathbb{R}^n)$ and $\|Mf\|_\infty \leqslant \|f\|_\infty$.

T H E M O R A L : 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 larger 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 2.10. 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 $f \in L^1_{\text{loc}}(\mathbb{R}^n)$.

Remark 2.11. $Mf \in L^1(\mathbb{R}^n)$ implies $f = 0$.

Reason. Let $r > 0$ and let $x \in \mathbb{R}^n$ such that $|x| \geqslant r$. Since $|y-x| \leqslant |y|+|x| < r+|x| \leqslant 2|x|$ whenever $|y| < r$, we conclude that $B(0,r) \subset B(x,2|x|)$. This implies

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

For a contradiction, assume that $f \neq 0$. 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 2.12. Let $f : \mathbb{R} \to \mathbb{R}$,

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

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

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

For $0 < x < \frac{1}{2}$, we have

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

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 weak L^1 space.

Definition 2.13. 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 2.14:

(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 $L^1(\mathbb{R}^n)$. ■

(3) The weak L^1 space is sometimes denoted by $L^{1,\infty}(\mathbb{R}^n)$ and it can be equipped with a seminorm

$$\|f\|_{L^{1,\infty}(\mathbb{R}^n)} = \sup_{\lambda>0} \lambda|\{x \in \mathbb{R}^n : |f(x)| > \lambda\}|.$$

The seminorm has the properties (exercise)

(a) $\|f\|_{L^{1,\infty}(\mathbb{R}^n)} = 0 \iff f(x) = 0$ for almost every $x \in \mathbb{R}^n$,

(b) $\|af\|_{L^{1,\infty}(\mathbb{R}^n)} = |a|\|f\|_{L^{1,\infty}(\mathbb{R}^n)}$ for every $a \in \mathbb{R}$ and

(c) $\|f+g\|_{L^{1,\infty}(\mathbb{R}^n)} \leqslant 2(\|f\|_{L^{1,\infty}(\mathbb{R}^n)} + \|g\|_{L^{1,\infty}(\mathbb{R}^n)})$.

Next we discuss an L^1 estimate for the Hardy-Littlewood maximal operator. By Remark 2.11, we recall that the Hardy-Littlewood maximal is integrable only if the function is zero almost everywhere. In particular, this shows that there is no hope to show that the Hardy-Littlewood maximal operator would be bounded on L^1. Our 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 2.15 (Covering lemma). Let $\mathscr{F}$ be a collection of open balls B such that

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

Then there is a countable (or finite) subcollection of pairwise disjoint balls $B(x_i,r_i) \in \mathscr{F}$, $i=1,2,\dots$, 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 $B(x_i,r_i)$, $i=1,2,\dots$, which covers the union of the original balls. In general, this is not possible, if we do not expand the balls. Thus

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

Note the measure of A can be estimated by the measure of the union of the balls and the measures of $\bigcup_{x\in A} B(x,r_x)$, $\bigcup_{i=1}^{\infty} B(x_i,5r_i)$ and $\bigcup_{i=1}^{\infty} B(x_i,r_i)$ are comparable.

THE STRATEGY OF 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 $B(x_1,r_1),\dots,B(x_{i-1},r_{i-1}) \in \mathscr{F}$ have been selected. Let

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

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 $B(x_1,r_1),\dots,B(x_{i-1},r_{i-1})$. Otherwise, we choose $B(x_i,r_i)\in\mathscr{F}$ such that

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

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

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

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

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

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

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. $\square$

Remarks 2.16:

(1) The factor 5 in the coverin lemma is not optimal. In fact, the same proof shows that this factor can be replaced with 3.

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

(3) The condition

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

in the covering lemma can be replaced by

$$\sup\{\operatorname{diam}(B):B\in\mathscr{F}\}<\infty,$$

see [4, Theorem 1, p. 27] and [9, Theorem 2.1].

(4) Some kind of boundedness assumption is needed in the covering lemma.

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 $\bigcup_{i=1}^{\infty} B(0,i) = \mathbb{R}^n$. ■

Next we discuss a weak type estimate for the Hardy-Littlewood maximal operator. One might want to apply Chebyshev's inequality and conclude that

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

However, this estimate is not useful, since Remark 2.11 shows that the Hardy-Littlewood maximal is integrable only if the function is zero almost everywhere. Thus the right-hand side is infinity unless the function is zero almost everywhere.

Theorem 2.17 (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

$$\frac{1}{|B(x,r_x)|}\int_{B(x,r_x)} |f(y)|\,dy > \lambda \tag{2.18}$$

We would like to apply the covering lemma, but the set $\bigcup_{x\in A_\lambda} B(x,r_x)$ is not necessarily bounded. To overcome this problem, we consider the sets $A_\lambda \cap B(0,k)$, $k = 1,2,\dots$. Let $\mathscr{F}$ be the collection of balls for which (2.18) 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

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

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

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

This implies

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

Finally,

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

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

Reason.

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

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. Fubini's theorem implies

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

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 $f_1 = f\chi_{\{|f|>\frac{\lambda}{2}\}}$, that is,

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

Then $|f_1(x)| > \frac{\lambda}{2}$ if $|f(x)| > \frac{\lambda}{2}$ and thus

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

This shows that $f_1 \in L^1(\mathbb{R}^n)$. On the other hand, $|f_2(x)| \leqslant \frac{\lambda}{2}$ for every $x \in \mathbb{R}^n$, which implies $\|f_2\|_\infty \leqslant \frac{\lambda}{2}$ 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 2.9, we have

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

From this we conclude using Lemma 2.7 that

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

for almost every $x \in \mathbb{R}^n$ and thus $Mf(x) > \lambda$ implies $Mf_1(x) > \frac{\lambda}{2}$ for almost every $x \in \mathbb{R}^n$. It follows from Theorem 2.17 that

$$\begin{aligned} |\{x \in \mathbb{R}^n : Mf(x) > \lambda\}| &\leqslant \left|\left\{x \in \mathbb{R}^n : Mf_1(x) > \tfrac{\lambda}{2}\right\}\right| \\ &\leqslant \frac{2\cdot 5^n}{\lambda} \|f_1\|_1 \\ &= \frac{2\cdot 5^n}{\lambda} \int_{\{x\in\mathbb{R}^n : |f(x)|>\frac{\lambda}{2}\}} |f(x)|\,dx \end{aligned}$$

for every $\lambda > 0$. By Cavalieri's principle and Fubini's theorem, as in the proof of Lemma 2.20, we obtain

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

This completes the proof. □

Remarks 2.23:

(1) The proof above gives

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

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 $p \to \infty$.

(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.

2.3 The Lebesgue differentiation theorem

The 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. This result holds at every point for a continuous function, see the proof of Lemma 2.4.

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 $x \in \mathbb{R}^n$ as $r \to 0$. ■

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 $x \in \mathbb{R}^n$. For a continuous function the inequality above holds at every point, see Lemma 2.4.

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 an 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.

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

(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 (2) and (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.

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

■

(6) If $f^*(x) > \lambda$, by (5) we have $Mf(x)+|f(x)| > \lambda$, from which we conclude that $Mf(x) > \frac{\lambda}{2}$ or $|f(x)| > \frac{\lambda}{2}$. By Theorem 2.17 and Chebyshev's inequality, we have

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

Finally, we are ready to prove the theorem. Recall from the measure and integration theory that compactly supported continuous functions are dense in $L^1(\mathbb{R}^n)$, see Theorem 1.57. 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

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

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 a 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 $L^1_{\mathrm{loc}}(\mathbb{R}^n)$.

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

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

In particular, the limit exists and it is independent of the representative in the equivalence class f. 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 $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$. Moreover, if f is a specific function in the equivalence class in $L^1_{\mathrm{loc}}(\mathbb{R}^n)$, 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) $A_i \subset B(x,r_i)$, $i = 1,2,\ldots$,

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

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

THE MORAL: 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

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

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

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

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

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

Thus $y \in B(x,r)$. ■

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

Reason.

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

■

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

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

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 $A = B(0,1)\setminus B(0,\frac{1}{2})$, then $A_r(x) = B(x,r)\setminus B(x,\frac{r}{2})$ 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.

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

□

Remark 2.32. The converse 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

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

2.4 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 let $F : [a,b] \to \mathbb{R}$,

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

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 $A_i = (x, x + r_i)$, $i = 1,2,\ldots$. 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 $A_i = (x - r_i, x)$, $i = 1,2,\ldots$, 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

$$F(x) = F(a) + \int_{[a,x]} F'(y)\,dy. \tag{2.35}$$

PROBLEM: What do we have to assume about the function F to guarantee that (2.35) holds?

(1) If $F \in C^1([a,b])$, then (2.35) holds.

(2) If $F = \chi_{[-1,1]}$ then $F' = 0$ almost everywhere in $\mathbb{R}$, but (2.35) 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 (2.35) 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 (2.35) 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 $x\in[a,b]$.

2.5 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 2.36. 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 2.37:

(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 2.38:

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

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

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

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

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. Moreover, a set $A \subset \mathbb{R}^n$ is measurable if and only if

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

For the proof, see [7, p. 464] and [9, Remarks 2.15 (2)].

2.6 The Sobolev embedding

This section discusses an application of the Hardy-Littlewood-Wiener theorem, see Theorem 2.22. 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,

$$u(x) = u(a) + \int_a^x u'(y)\,dy = \int_{-\infty}^x u'(y)\,dy, \tag{2.39}$$

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

$$u(x) = -\int_x^\infty u'(y)\,dy. \tag{2.40}$$

Equalities (2.39) and (2.40) 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 2.41. 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 $\omega_{n-1} = n\Omega_n$ is the $(n-1)$-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.

THE MORAL: This is a higher dimensional version of the fundamental theorem of calculus

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

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

□

By the Cauchy-Schwarz inequality and Lemma 2.41, we have

$$\begin{aligned}
|u(x)| &= \left|\frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{\nabla u(y)\cdot(x-y)}{|x-y|^n}\,dy\right|\\
&\leqslant \frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{|\nabla u(y)||x-y|}{|x-y|^n}\,dy\\
&= \frac{1}{\omega_{n-1}}\int_{\mathbb{R}^n} \frac{|\nabla u(y)|}{|x-y|^{n-1}}\,dy\\
&= \frac{1}{\omega_{n-1}} I_1(|\nabla u|)(x),
\end{aligned} \tag{2.42}$$

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

$$I_\alpha f(x) = \int_{\mathbb{R}^n} \frac{f(y)}{|x-y|^{n-\alpha}}\,dy. \tag{2.43}$$

Lemma 2.44. If $0<\alpha<n$, there exists a constant $c = c(n,\alpha) > 0$, such that

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

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

THE MORAL: Some other operator, in this case the Riesz potential, can be bounded by the maximal operator.

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

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

□

Theorem 2.45 (The Sobolev inequality for the Riesz potentials). Assume that $1 < p < n$ and $0 < \alpha < n/p$. Then there exists a constant $c = c(n,p,\alpha) > 0$, such that for every $f \in L^p(\mathbb{R}^n)$ we have

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

THE MORAL: The proof applies the Hardy-Littlewood-Wiener theorem to conclude a norm estimate for some other operator, in this case the Riesz potential.

Proof. If $f = 0$ almost everywhere, there is nothing to prove, and thus we may assume that $f > 0$ on a set of positive measure. This implies $Mf(x) > 0$ for every $x \in \mathbb{R}^n$. By Hölder's inequality

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

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

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

Lemma 2.44 implies

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

By choosing

$$r = \left(\frac{Mf(x)}{\|f\|_p}\right)^{-\frac{p}{n}} > 0,$$

we obtain

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

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

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

The Hardy-Littlewood theorem II (Theorem 2.22) implies

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

and thus

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

Corollary 2.46 (The Sobolev inequality). If $1 < p < n$, there exists a constant $c = c(n,p)$ such that

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

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

Proof. By (2.42), we have

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

Thus Theorem 2.45 implies

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

Remark 2.47. 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

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

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 Convolution

We begin with a formal definition of convolution.

Definition 3.1. 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

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y) g(x-y)\, dy,$$

whenever this makes sense.

THE MORAL: The convolution becomes a standard product of functions after taking the Fourier transform, see [7, Chapter 13].

WARNING: It is not clear that integral of the function $y \mapsto f(y)g(x-y)$ exists. This requires further analysis.

Remark 3.2. 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 $\Phi : \mathbb{R}^n \to \mathbb{R}^n$, $\Phi(y) = x - y$ is invertible and its inverse mapping 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. ∎

THE MORAL: 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 on $\mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$. This is needed in the application of Fubini's theorem, which ensures almost everywhere finiteness of $|f| * |g|$ under appropriate conditions.

Remarks 3.3:

(1) Assume that $f : \mathbb{R}^n \to [-\infty, \infty]$ is a Lebesgue measurable function on $\mathbb{R}^n$. Then $\widetilde{f} : \mathbb{R}^n \times \mathbb{R}^n \to [-\infty, \infty]$, $\widetilde{f}(x, y) = f(x - y)$ is a Lebesgue measurable function on $\mathbb{R}^{2n}$.

Reason. For an arbitrary set $E \subset \mathbb{R}^n$, let $\widetilde{E} = \{(x, y) : x - y \in E\}$. We show that $\{(x, y) : \widetilde{f}(x, y) < a\}$ is a Lebesgue measurable set in $\mathbb{R}^{2n}$ for every $a \in \mathbb{R}$. Let $A = \{z \in \mathbb{R}^n : f(z) < a\}$. Since f is a Lebesgue measurable function, the set $A \subset \mathbb{R}^n$ is Lebesgue measurable for every $a \in \mathbb{R}$. Then

$$\begin{aligned}\{(x, y) : \widetilde{f}(x, y) < a\} &= \{(x, y) : f(x - y) < a\} \\ &= \{(x, y) : x - y \in A\} = \widetilde{A}.\end{aligned}$$

Since A is a Lebesgue measurable set, there exists a G_δ set $G \supset A$ such that $G \setminus A$ is a set of n-dimensional Lebesgue measure zero. It follows that

$$\begin{aligned}\widetilde{E} &= \{(x, y) : x - y \in E\} = \{(x, y) : x - y \in G \setminus (G \setminus E)\} \\ &= \{(x, y) : x - y \in G\} \setminus \{(x, y) : x - y \in G \setminus E\} = \widetilde{G} \setminus \widetilde{G \setminus E}.\end{aligned}$$

We claim that $\widetilde{G}$ is a G_δ set and $\widetilde{G \setminus E}$ is a set of $2n$-dimensional Lebesgue measure zero. This implies that $\widetilde{E}$ is Lebesgue measurable in $\mathbb{R}^{2n}$. First we note that if $U \subset \mathbb{R}^n$ is an open set, then $\widetilde{U} \subset \mathbb{R}^{2n}$ is an open set. By considering countable intersections of open sets we see that if $G \subset \mathbb{R}^n$ is a G_δ set, then $\widetilde{G} \subset \mathbb{R}^{2n}$ is a G_δ set. Since $|G \setminus A| = 0$, there exist open sets $U_i \supset G \setminus A$ such that $|U_i| \to 0$ as $i \to \infty$. By a slight abuse of notation we denote both n-dimensional and $2n$-dimensional Lebesgue measures by $|\cdot|$. We compute $|\widetilde{U}_i \cap B(0, k)|$, $k = 1, 2, \dots$, by observing that

$$\chi_{\widetilde{U}_i \cap B(0,k)}(x, y) = \chi_{U_i}(x - y)\chi_{B(0,k)}(y)$$

for every $(x,y)\in\mathbb{R}^{2n}$. By Fubini's theorem we obtain

$$\begin{aligned}
|\widetilde{U}_i\cap B(0,k)| &= \int_{\mathbb{R}^{2n}} \chi_{\widetilde{U}_i\cap B(0,k)}(x,y)\,dx\,dy\\
&= \int_{\mathbb{R}^{2n}} \chi_{U_i}(x-y)\chi_{B(0,k)}(y)\,dx\,dy\\
&= \int_{\mathbb{R}^{n}} \left(\int_{\mathbb{R}^{n}} \chi_{U_i}(x-y)\,dx\right)\chi_{B(0,k)}(y)\,dy\\
&= \int_{\mathbb{R}^{n}} \left(\int_{\mathbb{R}^{n}} \chi_{U_i}(x)\,dx\right)\chi_{B(0,k)}(y)\,dy\\
&= |U_i||B(0,k)|, \quad k=1,2,\dots.
\end{aligned}$$

Here we also applied the translation invariance of the Lebesgue integral. Since $\widetilde{G\setminus A}\cap B(0,k)\subset\widetilde{U}_i\cap B(0,k)$ for every $i,k=1,2,\dots$ and $|U_i|\to 0$ as $i\to\infty$, we conclude that

$$|\widetilde{G\setminus A}\cap B(0,k)|\leqslant|\widetilde{U}_i\cap B(0,k)|=|U_i||B(0,k)|\to 0$$

as $i\to\infty$ and thus $|\widetilde{G\setminus A}\cap B(0,k)|=0$ for every $k=1,2,\dots$. Finally, we note that

$$|\widetilde{G\setminus A}|=\left|\bigcup_{k=1}^{\infty}\widetilde{G\setminus A}\cap B(0,k)\right|\leqslant\sum_{k=1}^{\infty}|\widetilde{G\setminus A}\cap B(0,k)|=0.$$

■

(2) Assume that $f:\mathbb{R}^n\to[-\infty,\infty]$ is a Lebesgue measurable function on $\mathbb{R}^n$. Then $\widetilde{f}:\mathbb{R}^{2n}\to[-\infty,\infty]$ $\widetilde{f}(x,y)=f(y)$ is a Lebesgue measurable measurable function on $\mathbb{R}^{2n}$.

Reason. Since f is a Lebesgue measurable function, the set $A=\{y\in\mathbb{R}^n: f(y)<a\}$ is Lebesgue measurable in $\mathbb{R}^n$ for every $a\in\mathbb{R}$. Since

$$\{(x,y)\in\mathbb{R}^{2n}:\widetilde{f}(x,y)<a\}=\mathbb{R}^n\times A,$$

we conclude that the set is Lebesgue measurable for every $a\in\mathbb{R}$. Thus $\widetilde{f}$ is a Lebesgue measurable function on $\mathbb{R}^{2n}$. ■

(3) Assume that $f,g:\mathbb{R}^n\to[-\infty,\infty]$ are Lebesgue measurable functions on $\mathbb{R}^n$. Then $\widetilde{f}:\mathbb{R}^{2n}\to[-\infty,\infty]$ $\widetilde{f}(x,y)=f(y)g(x-y)$ is a Lebesgue measurable measurable function on $\mathbb{R}^{2n}$.

Reason. The function $(x,y)\mapsto f(y)$ is Lebesgue measurable by (2) and the function $(x,y)\mapsto g(x-y)$ is Lebesgue measurable by (1). As a product of two measurable functions, the function $(x,y)\mapsto f(y)g(x-y)$ is Lebesgue measurable. ■

The next result settles the integrability questions in the definition of the convolution under certain assumptions.

Theorem 3.4 (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 and translation invariance of Lebesgue integral, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} (f * g)(x)\,dx &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y)g(x-y)\,dy\,dx \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(y)g(x-y)\,dx\,dy \\
&= \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} g(x-y)\,dx \right) dy \\
&= \int_{\mathbb{R}^n} f(y) \left(\int_{\mathbb{R}^n} g(x)\,dx \right) dy \\
&= \int_{\mathbb{R}^n} f(y)\,dy \int_{\mathbb{R}^n} g(x)\,dx.
\end{aligned}$$

Thus $\|f * g\|_1 = \|f\|_1 \|g\|_1$ 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

$$\begin{aligned}|(f*g)(x)| &\leqslant \int_{\mathbb{R}^n} |f(y)||g(x-y)|\,dy \\ &= \int_{\mathbb{R}^n} |f(y)||g(x-y)|^{\frac{1}{p}}|g(x-y)|^{\frac{1}{p'}}\,dy \\ &\leqslant \left(\int_{\mathbb{R}^n} |f(y)|^p|g(x-y)|\,dy\right)^{\frac{1}{p}}\left(\int_{\mathbb{R}^n} |g(x-y)|\,dy\right)^{\frac{1}{p'}} \\ &= \left(\int_{\mathbb{R}^n} |f(y)|^p|g(x-y)|\,dy\right)^{\frac{1}{p}}\left(\int_{\mathbb{R}^n} |g(y)|\,dy\right)^{\frac{1}{p'}}.\end{aligned}$$

This implies that

$$|(f*g)(x)|^p \leqslant \|g\|_1^{\frac{p}{p'}} \int_{\mathbb{R}^n} |f(y)|^p|g(x-y)|\,dy$$

and by Fubini's theorem we have

$$\begin{aligned}\int_{\mathbb{R}^n} |(f*g)(x)|^p\,dx &\leqslant \|g\|_1^{\frac{p}{p'}} \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(y)|^p|g(x-y)|\,dy\,dx \\ &= \|g\|_1^{\frac{p}{p'}} \int_{\mathbb{R}^n} |f(y)|^p\left(\int_{\mathbb{R}^n} |g(x-y)|\,dx\right)dy \\ &= \|g\|_1^{\frac{p}{p'}} \int_{\mathbb{R}^n} |f(y)|^p\left(\int_{\mathbb{R}^n} |g(y)|\,dx\right)dy \\ &= \|g\|_1^{\frac{p}{p'}}\|g\|_1\|f\|_p^p = \|g\|_1^p\|f\|_p^p.\end{aligned}$$

□

Remark 3.5. 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 $x = 0$.

The following lemma shows that the convolution regarded as a multiplication in $L^1(\mathbb{R}^n)$ satisfies certain standard algebraic laws.

Lemma 3.6. 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) □

Theorem 3.7. 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 that $1 \leqslant p < \infty$. By Hölder's inequality and translation invariance of Lebesgue integral, we have

$$\begin{aligned}
|(f*g)(x)| &= \left| \int_{\mathbb{R}^n} f(x-y)g(y)\,dy \right| \\
&\leqslant \left(\int_{\mathbb{R}^n} |f(x-y)|^p\,dy \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \left(\int_{\mathbb{R}^n} |f(y)|^p\,dy \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \|f\|_p \|g\|_{p'} < \infty
\end{aligned}$$

for every $x \in \mathbb{R}^n$. This implies that $\|f*g\|_\infty \leqslant \|f\|_p \|g\|_{p'}$.

By Hölder's inequality, and by reflection and translation invariances of Lebesgue integral, we have

$$\begin{aligned}
|(f*g)(x)-(f*g)(z)| &= \left| \int_{\mathbb{R}^n} (f(x-y)-f(z-y))g(y)\,dy \right| \\
&\leqslant \left(\int_{\mathbb{R}^n} |f(x-y)-f(z-y)|^p\,dy \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \left(\int_{\mathbb{R}^n} |f(y-x)-f(y-z)|^p\,dy \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \left(\int_{\mathbb{R}^n} |f(y-x)-f(y-z)|^p\,dy \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \left(\int_{\mathbb{R}^n} |f(v+z-x)-f(v)|^p\,dv \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g(y)|^{p'}\,dy \right)^{\frac{1}{p'}} \\
&= \|\tau_{z-x}f - f\|_p \|g\|_{p'}.
\end{aligned}$$

By Theorem 1.61, 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 $x, z \in \mathbb{R}^n$ with $|z-x| < \delta$. This shows that $f*g$ is uniformly continuous. □

3.2 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 3.7 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 3.8. Assume that $\phi \in L^1(\mathbb{R}^n)$. For $\varepsilon > 0$, let

$$\phi_\varepsilon(x) = \frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right), \quad x \in \mathbb{R}^n.$$

Such a collection of functions is called an approximation of the identity.

Remark 3.9. Let $\phi \in C_0(\mathbb{R}^n)$.

(1) Since a continuous function attains its maximum value in a compact set, we have

$$\sup_{x\in\mathbb{R}^n} |\phi(x)| = \max_{x\in\mathbb{R}^n} |\phi(x)| < \infty.$$

The definition of ϕ_ε implies

$$\|\phi_\varepsilon\|_\infty = \max_{x\in\mathbb{R}^n} |\phi_\varepsilon(x)| = \frac{1}{\varepsilon^n}\max_{x\in\mathbb{R}^n} |\phi(x)| = \frac{1}{\varepsilon^n}\|\phi\|_\infty, \quad \varepsilon > 0.$$

Unless $\phi(x) = 0$ for every $x \in \mathbb{R}^n$, we have

$$\lim_{\varepsilon\to 0}\max_{x\in\mathbb{R}^n} |\phi_\varepsilon(x)| = \infty.$$

For the supports we have

$$\operatorname{supp}\phi_\varepsilon = \varepsilon \operatorname{supp}\phi, \quad \varepsilon > 0.$$

Reason. Since

$$\begin{aligned}\{x \in \mathbb{R}^n : \phi_\varepsilon(x) \neq 0\} &= \left\{x \in \mathbb{R}^n : \frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right) \neq 0\right\} = \left\{x \in \mathbb{R}^n : \phi\left(\frac{x}{\varepsilon}\right) \neq 0\right\} \\ &= \left\{\varepsilon x \in \mathbb{R}^n : \phi(x) \neq 0\right\} = \varepsilon\{x \in \mathbb{R}^n : \phi(x) \neq 0\},\end{aligned}$$

we have

$$\begin{aligned}\operatorname{supp}\phi_\varepsilon &= \overline{\{x \in \mathbb{R}^n : \phi_\varepsilon(x) \neq 0\}} = \overline{\varepsilon\{x \in \mathbb{R}^n : \phi(x) \neq 0\}} \\ &= \varepsilon\overline{\{x \in \mathbb{R}^n : \phi(x) \neq 0\}} = \varepsilon \operatorname{supp}\phi.\end{aligned}$$

■

Since $\operatorname{supp}\phi$ is compact and $\operatorname{supp}\phi_\varepsilon$ is compact for every $\varepsilon > 0$. Thus $|\operatorname{supp}\phi| < \infty$ and

$$\lim_{\varepsilon\to 0}|\operatorname{supp}\phi_\varepsilon| = \lim_{\varepsilon\to 0}|\varepsilon \operatorname{supp}\phi| = \lim_{\varepsilon\to 0}\varepsilon^n|\operatorname{supp}\phi| = 0.$$

(2) Let $1 \leqslant p < \infty$. By the change of variables $y = \frac{x}{\varepsilon}$, $dx = \varepsilon^n\,dy$, we have

$$\begin{aligned}\int_{\mathbb{R}^n}|\phi_\varepsilon(x)|^p\,dx &= \int_{\mathbb{R}^n}\left|\frac{1}{\varepsilon^n}\phi\left(\frac{x}{\varepsilon}\right)\right|^p dx = \frac{1}{\varepsilon^{np}}\int_{\mathbb{R}^n}\left|\phi\left(\frac{x}{\varepsilon}\right)\right|^p dx \\ &= \frac{1}{\varepsilon^{np}}\varepsilon^n\int_{\mathbb{R}^n}|\phi(y)|^p\,dy = \varepsilon^{n(1-p)}\int_{\mathbb{R}^n}|\phi(y)|^p\,dy\end{aligned}$$

and thus

$$\|\phi_\varepsilon\|_p = \left(\frac{1}{\varepsilon^n}\right)^{\frac{p-1}{p}}\|\phi\|_p.$$

THE MORAL: Smaller values of $\varepsilon > 0$ produce higher peaks and smaller supports. Convolution with approximations of the identity is expected to act as the identity operator on a class of functions as $\varepsilon \to 0$.

Example 3.10. Let $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$\phi(x) = \frac{\chi_{B(0,1)}(x)}{|B(0,1)|}.$$

Then

$$\phi_\varepsilon(x) = \frac{1}{\varepsilon^n}\frac{\chi_{B(0,1)}(\frac{x}{\varepsilon})}{|B(0,1)|} = \frac{\chi_{B(0,\varepsilon)}(x)}{|B(0,\varepsilon)|}, \quad \varepsilon > 0.$$

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 theorem (Theorem 3.4)

$$\|f * \phi_\varepsilon\|_1 \leqslant \|f\|_1\|\phi_\varepsilon\|_1 = \|f\|_1 \quad \text{for every} \quad \varepsilon > 0,$$

since $\|\phi_\varepsilon\|_1 = 1$ for every $\varepsilon > 0$. By the Lebesgue differentiation theorem (Theorem 2.24) we have

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

for almost every $x \in \mathbb{R}^n$. Observe, that we have

$$|(f * \phi_\varepsilon)(x)| = \left|\frac{1}{|B(x,\varepsilon)|}\int_{B(x,\varepsilon)} f(y)\,dy\right| \leqslant \frac{1}{|B(x,\varepsilon)|}\int_{B(x,\varepsilon)}|f(y)|\,dy \leqslant Mf(x)$$

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

$$\sup_{\varepsilon>0}|(f * \phi_\varepsilon)(x)| \leqslant Mf(x)$$

for every $x \in \mathbb{R}^n$. This kind of bound for a more general mollifier ϕ is discussed in Theorem 3.13.

We assumed that $\phi \in L^\infty(\mathbb{R}^n)$ and $\operatorname{supp}\phi = \overline{B(0,1)}$ in the previous example. Next we discuss properties of a general approximation of the identity with $\phi \in L^1(\mathbb{R}^n)$.

Lemma 3.11. 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 assertion (1) explains the scaling factors in the definition of ϕ_ε. These are chosen so that the integral of ϕ_ε is independent of $\varepsilon > 0$. The assertion (2) tells that the integral of $|\phi_\varepsilon|$ becomes as small as we please for $\varepsilon > 0$ small enough. This indicates that $|\phi_\varepsilon|$ concentrated in a small neighbourhood of the point x. Note that assertions (1) and (2) hold for compactly supported continuous functions.

Remark 3.12. The assertion (2) is clear, if the support of ϕ is a compact set.

Proof. (1) By a change of variables $y = \frac{x}{\varepsilon}$, $dx = \varepsilon^n\,dy$, we have

$$\int_{\mathbb{R}^n} \phi_\varepsilon(x)\,dx = \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} \phi\left(\frac{x}{\varepsilon}\right)dx = \int_{\mathbb{R}^n} \phi(y)\,dy.$$

(2) By the same change of variables as above

$$\begin{aligned}\int_{\mathbb{R}^n\setminus B(0,r)} |\phi_\varepsilon(x)|\,dx &= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n\setminus B(0,r)} \left|\phi\left(\frac{x}{\varepsilon}\right)\right| dx \\ &= \int_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})} |\phi(y)|\,dy \\ &= \int_{\mathbb{R}^n} |\phi(x)|\chi_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})}(x)\,dx \xrightarrow{\varepsilon\to 0} 0\end{aligned}$$

by the dominated convergence theorem with the integrable dominating function

$$|\phi|\chi_{\mathbb{R}^n\setminus B(0,\frac{r}{\varepsilon})} \leqslant |\phi| \in L^1(\mathbb{R}^n) \quad \text{for every} \quad \varepsilon > 0. \qquad \square$$

3.3 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 $\phi(x) \leqslant \phi(y)$. The next result generalizes the bound in Example 3.10 for a more general mollifier ϕ.

Theorem 3.13. 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 pointwise upper bound for many other operators as well. In this sense the maximal function controls the weighted 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

$$\phi = \sum_i a_i \chi_{B(0,r_i)}$$

with $a_i > 0$. The sum here is over finitely many indices only. Then

$$\|\phi\|_1 = \int_{\mathbb{R}^n} \sum_i a_i \chi_{B(0,r_i)}\, dx = \sum_i \int_{\mathbb{R}^n} a_i \chi_{B(0,r_i)}\, dx = \sum_i a_i |B(0,r_i)|.$$

By a change of variables $z = \frac{y}{\varepsilon}$, $y = \varepsilon z$, $dy = \varepsilon^n\, dz$, we have

$$\begin{aligned}
|(\phi_\varepsilon * f)(x)| &= \left| \int_{\mathbb{R}^n} f(x-y)\phi_\varepsilon(y)\, dy \right| = \left| \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} f(x-y)\phi\Big(\frac{y}{\varepsilon}\Big)\, dy \right| \\
&= \left| \int_{\mathbb{R}^n} f(x-\varepsilon z)\phi(z)\, dz \right| = \left| \sum_i \int_{B(0,r_i)} f(x-\varepsilon z) a_i\, dz \right| \\
&\leqslant \sum_i a_i \int_{B(0,r_i)} |f(x-\varepsilon z)|\, dz \\
&= \sum_i a_i |B(0,r_i)| \frac{1}{|B(0,r_i)|} \int_{B(0,r_i)} |f(x-\varepsilon z)|\, dz.
\end{aligned}$$

Again, by a change of variables $y = x - \varepsilon z$, $z = \frac{1}{\varepsilon}(y-x)$, $z = \varepsilon^{-n}\, dy$, we have

$$\begin{aligned}
\frac{1}{|B(0,r_i)|} \int_{B(0,r_i)} |f(x-\varepsilon z)|\, dz &= \frac{1}{\varepsilon^n |B(0,r_i)|} \int_{B(x,\varepsilon r_i)} |f(y)|\, dy \\
&= \frac{1}{|B(x,\varepsilon r_i)|} \int_{B(x,\varepsilon r_i)} |f(y)|\, dy \leqslant Mf(x).
\end{aligned}$$

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 nonnegative 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 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 $x \in \mathbb{R}^n$. □

The next result generalizes the pointwise convergence result in Example 3.10 for more general mollifiers.

Theorem 3.14. Assume that $\phi \in L^1(\mathbb{R}^n)$ is nonnegative, radial and decreasing and let $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p \leqslant \infty$. Then

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * f)(x) = \|\phi\|_1 f(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

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 approximation of the identity by

$$M_\phi f(x) = \sup_{\varepsilon>0} |(\phi_\varepsilon * f)(x)|.$$

By Theorem 3.13

$$M_\phi f(x) \leqslant \|\phi\|_1 Mf(x) \quad \text{for every} \quad x \in \mathbb{R}^n.$$

By the weak type estimate for the Hardy-Littlewood maximal function, see Theorem 2.17, for $f \in L^1(\mathbb{R}^n)$, we have

$$|\{x \in \mathbb{R}^n : M_\phi f(x) > \lambda\}| \leqslant |\{x \in \mathbb{R}^n : \|\phi\|_1 Mf(x) > \lambda\}| \leqslant \frac{5^n \|\phi\|_1}{\lambda} \|f\|_1$$

for every $\lambda > 0$.

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

$$\begin{aligned}|\{x \in \mathbb{R}^n : M_\phi f(x) > \lambda\}| &\leqslant |\{x \in \mathbb{R}^n : \|\phi\|_1 Mf(x) > \lambda\}| \\ &\leqslant \frac{\|\phi\|_1^p}{\lambda^p} \|Mf\|_p^p \\ &\leqslant c(n,p)\frac{\|\phi\|_1^p}{\lambda^p} \|f\|_p^p\end{aligned} \tag{3.15}$$

for every $\lambda > 0$. Thus (3.15) holds for $1 \leqslant p < \infty$.

The proof of the claim is based on these two estimates and is somewhat similar to the proof of Theorem 2.24. Let $a = \|\phi\|_1$ and $\eta > 0$. 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 $|y| < \delta$. By Lemma 3.11 (1), we have

$$ag(x) = g(x)\int_{\mathbb{R}^n}\phi(y)\,dy = g(x)\int_{\mathbb{R}^n}\phi_\varepsilon(y)\,dy = \int_{\mathbb{R}^n} g(x)\phi_\varepsilon(y)\,dy.$$

This implies

$$\begin{aligned}
|(\phi_\varepsilon * g)(x) - ag(x)| &= \left|\int_{\mathbb{R}^n} g(x-y)\phi_\varepsilon(y)\,dy - \int_{\mathbb{R}^n} g(x)\phi_\varepsilon(y)\,dy\right| \\
&= \left|\int_{\mathbb{R}^n} (g(x-y)-g(x))\phi_\varepsilon(y)\,dy\right| \\
&\leqslant \int_{\mathbb{R}^n} |g(x-y)-g(x)|\phi_\varepsilon(y)\,dy \\
&= \int_{B(0,\delta)} \dots\,dy + \int_{\mathbb{R}^n\setminus B(0,\delta)} \dots\,dy \\
&\leqslant \eta\int_{B(0,\delta)}\phi_\varepsilon(y)\,dy + 2\|g\|_\infty \int_{\mathbb{R}^n\setminus B(0,\delta)}\phi_\varepsilon(y)\,dy.
\end{aligned}$$

By Lemma 3.11

$$\int_{B(0,\delta)}\phi_\varepsilon(y)\,dy \leqslant \|\phi_\varepsilon\|_1 = \|\phi\|_1$$

and

$$\int_{\mathbb{R}^n\setminus B(0,\delta)}\phi_\varepsilon(y)\,dy \xrightarrow{\varepsilon\to 0} 0.$$

By letting first $\eta \to 0$ and then $\varepsilon \to 0$, we have

$$\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * g)(x) - ag(x)| = 0 \quad \text{for every} \quad x \in \mathbb{R}^n.$$

This shows that

$$\lim_{\varepsilon\to 0}(\phi_\varepsilon * g)(x) = ag(x) \quad \text{for every} \quad x \in \mathbb{R}^n,$$

that is, the claim of the theorem holds for $g \in C_0(\mathbb{R}^n)$ at every point.

Then we consider the corresponding claim for $f \in L^p(\mathbb{R}^n)$. We note that

$$\begin{aligned}
&\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| \\
&\quad\leqslant \limsup_{\varepsilon\to 0} |\phi_\varepsilon * (f-g)(x) - a(f-g)(x)| + \underbrace{\limsup_{\varepsilon\to 0} |(\phi_\varepsilon * g)(x) - ag(x)|}_{=0} \\
&\quad\leqslant M_\phi(f-g)(x) + a|(f-g)(x)|.
\end{aligned}$$

Let

$$A_i = \left\{x \in \mathbb{R}^n : \limsup_{\varepsilon\to 0} |(\phi_\varepsilon * f)(x) - af(x)| > \tfrac{1}{i}\right\}, \quad i = 1,2,\dots.$$

As in the proof of Theorem 2.24

$$A_i \subset \left\{x \in \mathbb{R}^n : M_\phi(f-g)(x) > \tfrac{1}{2i}\right\} \cup \left\{x \in \mathbb{R}^n : |f(x)-g(x)| > \tfrac{1}{2i}\right\}, \quad i = 1,2,\dots,$$

and by (3.15) and Chebyshev's inequality we have

$$\begin{aligned}|A_i| &\leqslant \left|\left\{x \in \mathbb{R}^n : M_\phi(f-g)(x) > \tfrac{1}{2i}\right\}\right| + \left|\left\{x \in \mathbb{R}^n : |f(x)-g(x)| > \tfrac{1}{2i}\right\}\right| \\ &\leqslant c i^p \|f-g\|_p^p + (2i)^p \|f-g\|_p^p \\ &= c i^p \|f-g\|_p^p \leqslant c i^p \eta^p, \quad i = 1,2,\dots,\end{aligned}$$

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

$$\begin{aligned}|(\phi_\varepsilon * f_2)(x)| &= \left|\int_{\mathbb{R}^n} f_2(x-y)\phi_\varepsilon(y)\,dy\right| \\ &= \left|\int_{\mathbb{R}^n\setminus B(0,1)} f_2(x-y)\phi_\varepsilon(y)\,dy\right| \\ &= \|f_2\|_\infty \int_{\mathbb{R}^n\setminus B(0,1)} \phi_\varepsilon(y)\,dy \xrightarrow{\varepsilon\to 0} 0.\end{aligned}$$

□

Remark 3.16. Under the assumptions of Theorem 3.14, if $f \in L^\infty(\mathbb{R}^n)$ is continuous at x, then

$$\lim_{\varepsilon\to 0} (\phi_\varepsilon * f)(x) = \|\phi\|_1 f(x).$$

Moreover, if $f \in L^\infty(\mathbb{R}^n)$ is uniformly continuous, the convergence is uniform. (Exercise)

Remark 3.17. Assume that the assumptions of Theorem 3.14 hold and let $f \in L^p(\mathbb{R}^n)$, $1 < p < \infty$. Let $a = \|\phi\|_1$. By Theorem 3.14

$$\lim_{\varepsilon \to 0} (\phi_\varepsilon * f)(x) = af(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

By Theorem 3.13

$$\sup_{\varepsilon > 0} |(\phi_\varepsilon * f)(x)| \leqslant aMf(x)$$

for every $x \in \mathbb{R}^n$. Theorem 2.22 implies that $Mf \in L^p(\mathbb{R}^n)$. This shows that

$$\begin{aligned} \sup_{\varepsilon > 0} |(\phi_\varepsilon * f)(x) - af(x)| &\leqslant \sup_{\varepsilon > 0} |(\phi_\varepsilon * f)(x)| + a|f(x)| \\ &\leqslant aMf(x) + a|f(x)| \end{aligned}$$

for almost every $x \in \mathbb{R}^n$ with $a(Mf + |f|) \in L^p(\mathbb{R}^n)$. Thus we may apply the dominated convergence theorem to conclude

$$\lim_{\varepsilon \to 0} \int_{\mathbb{R}^n} |(\phi_\varepsilon * f)(x) - af(x)|^p \, dx = \int_{\mathbb{R}^n} \lim_{\varepsilon \to 0} |(\phi_\varepsilon * f)(x) - af(x)|^p \, dx = 0.$$

This shows that

$$\lim_{\varepsilon \to 0} \|\phi_\varepsilon * f - af\|_p = 0.$$

THE MORAL: In this case almost everywhere pointwise convergence upgrades to L^p convergence by the Hardy-Littlewood maximal function theorem for $1 < p < \infty$. However, this argument does not work for $p = 1$, since the Hardy-Littlewood maximal operator is not bounded on L^1. Theorem 3.18 below gives a general result that applies for $1 \leqslant p < \infty$ and for a general mollifier $\phi \in L^1(\mathbb{R}^n)$.

3.4 Convergence in L^p

Theorem 3.14 asserts that a convolution approximation of a L^p function 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 for approximations of the identity.

Theorem 3.18. 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: Approximations of the identity converge in L^p for $1 \leqslant p < \infty$.

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,

$$\lim_{\varepsilon \to 0} \|\phi_\varepsilon * f - af\|_\infty = 0.$$

Proof. By Lemma 3.11 (1), we have

$$af(x) = f(x)\int_{\mathbb{R}^n} \phi(y)\,dy = f(x)\int_{\mathbb{R}^n} \phi_\varepsilon(y)\,dy = \int_{\mathbb{R}^n} f(x)\phi_\varepsilon(y)\,dy.$$

$\boxed{p=1}$ We note that

$$\begin{aligned}|(f*\phi_\varepsilon)(x) - af(x)| &= \left|\int_{\mathbb{R}^n} (f(x-y)-f(x))\phi_\varepsilon(y)\,dy\right| \\ &\leqslant \int_{\mathbb{R}^n} |f(x-y)-f(x)||\phi_\varepsilon(y)|\,dy.\end{aligned}$$

By Fubini's theorem

$$\begin{aligned}&\int_{\mathbb{R}^n} \left|(f*\phi_\varepsilon)(x) - af(x)\right| dx \\ &\quad\leqslant \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(x-y)-f(x)||\phi_\varepsilon(y)|\,dy\,dx \\ &\quad= \int_{\mathbb{R}^n} |\phi_\varepsilon(y)| \left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|\,dx\right) dy.\end{aligned}$$

$\boxed{1<p<\infty}$ We note that

$$\begin{aligned}|(f*\phi_\varepsilon)(x) - af(x)| &= \left|\int_{\mathbb{R}^n} (f(x-y)-f(x))\phi_\varepsilon(y)\,dy\right| \\ &\leqslant \int_{\mathbb{R}^n} |f(x-y)-f(x)||\phi_\varepsilon(y)|^{\frac{1}{p}}|\phi_\varepsilon(y)|^{\frac{1}{p'}}\,dy,\end{aligned}$$

where $\frac{1}{p}+\frac{1}{p'}=1$. By Hölder's inequality and Fubini's theorem

$$\begin{aligned}&\int_{\mathbb{R}^n} \left|(f*\phi_\varepsilon)(x) - af(x)\right|^p dx \\ &\quad\leqslant \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|^p|\phi_\varepsilon(y)|\,dy\right)\left(\int_{\mathbb{R}^n} |\phi_\varepsilon(y)| dy\right)^{\frac{p}{p'}} dx \\ &\quad= \|\phi\|_1^{\frac{p}{p'}} \int_{\mathbb{R}^n} |\phi_\varepsilon(y)| \left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|^p\,dx\right) dy.\end{aligned}$$

For $1\leqslant p<\infty$, we have

$$\int_{\mathbb{R}^n} |\phi_\varepsilon(y)| \left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|^p\,dx\right) dy = \int_{B(0,r)} \dots dy + \int_{\mathbb{R}^n\setminus B(0,r)} \dots dy$$

Let $\eta>0$. By Theorem 1.61 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

$$\begin{aligned}&\|\phi\|_1^{\frac{p}{p'}} \int_{B(0,r)} |\phi_\varepsilon(y)| \left(\int_{\mathbb{R}^n} |f(x-y)-f(x)|^p\,dx\right) dy \\ &\quad\leqslant \frac{\eta}{2}\frac{\|\phi\|_1^{\frac{p}{p'}+1}}{\|\phi\|_1^p+1} \leqslant \frac{\eta}{2},\end{aligned}$$

where we note that $\frac{p}{p'}+1=p$.

By Lemma 3.11 (2), there exists $\varepsilon'>0$ such that

$$\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\,dy<\frac{\eta}{2^{p+2}(\|f\|_p^p\|\phi\|_1^{\frac{p}{p'}}+1)},\quad\text{whenever}\quad 0<\varepsilon<\varepsilon'.$$

This shows that

$$\begin{aligned}
&\|\phi\|_1^{\frac{p'}{p}}\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}|f(x-y)-f(x)|^p\,dx\right)dy\\
&\leqslant\|\phi\|_1^{\frac{p'}{p}}\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}2^p(|f(x-y)|^p+|f(x)|^p)\,dx\right)dy\\
&\leqslant 2^p\|\phi\|_1^{\frac{p'}{p}}\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\left(\int_{\mathbb{R}^n}(|f(x)|^p+|f(x)|^p)\,dx\right)dy\\
&\leqslant 2^{p+1}\|\phi\|_1^{\frac{p}{p'}}\|f\|_p^p\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_\varepsilon(y)|\,dy\\
&<\frac{\eta}{2},\quad\text{whenever}\quad 0<\varepsilon<\varepsilon'.
\end{aligned}$$

Thus

$$\|f*\phi_\varepsilon-af\|_p^p<\frac{\eta}{2}+\frac{\eta}{2}=\eta.$$

□

Remark 3.19. An examination of the proof above shows that a more general result holds as well. Let $\phi_i\in L^1(\mathbb{R}^n)$, $i=1,2,\dots$, be a sequence with the properties

(1) $\lim_{i\to\infty}\int_{\mathbb{R}^n}\phi_i(x)\,dx=a$,

(2) $\sup_i\int_{\mathbb{R}^n}|\phi_i(x)|\,dx<\infty$ and

(3) $\lim_{i\to\infty}\int_{\mathbb{R}^n\setminus B(0,r)}|\phi_i(x)|\,dx=0$ for every $r>0$.

Then

$$\lim_{i\to\infty}\|\phi_i*f-af\|_p=0.$$

Note that here ϕ_i do not have to be nonnegative or given by the formula for the approximate identity.

3.5 Smoothing in the entire space

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. The next example shows that there exist such functions.

Example 3.20. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi(x) = \begin{cases} \exp\left(\frac{1}{|x|^2-1}\right), & |x| < 1, \\ 0, & |x| \geqslant 1. \end{cases}$$

Let ϕ_ε, $\varepsilon > 0$, be an approximation of the identity as in Definition 3.8. 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 $\operatorname{supp}\phi_\varepsilon = \overline{B(0,\varepsilon)}$. By a change of variables $y = \frac{x}{\varepsilon}$, $dx = \varepsilon^n\,dy$, we have

$$\begin{aligned}\int_{\mathbb{R}^n} \phi_\varepsilon(x)\,dx &= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} \phi\left(\frac{x}{\varepsilon}\right)dx \\ &= \frac{1}{\varepsilon^n}\int_{\mathbb{R}^n} \phi(y)\varepsilon^n\,dy = \int_{\mathbb{R}^n} \phi(y)\,dy \\ &= \int_{\mathbb{R}^n} \frac{\varphi(x)}{\|\varphi\|_1}\,dx = \frac{\|\varphi\|_1}{\|\phi\|_1} = 1, \quad \varepsilon > 0.\end{aligned}$$

Young's theorem (Theorem 3.4) implies that

$$\|f * \phi_\varepsilon\|_1 \leqslant \|f\|_1 \|\phi_\varepsilon\|_1 = \|f\|_1 \quad \text{for every} \quad \varepsilon > 0.$$

The function ϕ_ε is called the standard mollifier. The function ϕ is not only continuous, but is is a compactly supported smooth function, that is, $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\operatorname{supp}\phi = \overline{B(0,1)}$ (exercise). In particular, this implies that $\phi_\varepsilon \in C_0^\infty(\mathbb{R}^n)$ with $\operatorname{supp}\phi = \overline{B(0,\varepsilon)}$, $\varepsilon > 0$. Hint: Let $h : \mathbb{R} \to \mathbb{R}$,

$$h(t) = \begin{cases} 0, & t \leqslant 0, \\ \exp\left(-\frac{1}{t}\right), & t > 0. \end{cases}$$

Then $h \in C^\infty(\mathbb{R})$. Prove by induction that $h^{(m)}(t) = P_m(\frac{1}{t})\exp(-\frac{1}{t})$ for $t > 0$, where P_m is a polynomial of degree $2m$. Then prove by induction that $h^{(m)}(0) = 0$. Then $\phi(x) = h(1 - |x|^2)$ 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 3.21. 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: Convolution inherits smoothness of the mollifier, since we differentiate under the integral sign. This is justified by the Lebesgue dominated convergence 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 3.22. Theorem 1.57 asserts that $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$. Theorem 3.21 and Theorem 3.18 imply that $C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$.

Proof. Theorem 3.7 implies that $f * \phi_\varepsilon$ is continuous. Let $e_i = (0, \dots, 1, \dots, 0)$, $i = 1, \dots, n$, be the standard ith basis vector in $\mathbb{R}^n$ and let $h \in \mathbb{R}$ with $0 < |h| < 1$. Then

$$\frac{(f * \phi_\varepsilon)(x + he_i) - (f * \phi_\varepsilon)(x)}{h} = \frac{1}{\varepsilon^n} \int_{\mathbb{R}^n} \frac{1}{h} \left[\phi\left(\frac{x + he_i - y}{\varepsilon} \right) - \phi\left(\frac{x - y}{\varepsilon} \right) \right] f(y)\, dy,$$

$i = 1, \dots, n$.

CLAIM:
$$\frac{1}{h} \left[\phi\left(\frac{x + he_i - y}{\varepsilon} \right) - \phi\left(\frac{x - y}{\varepsilon} \right) \right] \xrightarrow{h \to 0} \frac{1}{\varepsilon} \frac{\partial \phi}{\partial x_i} \left(\frac{x - y}{\varepsilon} \right).$$

Reason. Let
$$\varphi(x) = \phi\left(\frac{x - y}{\varepsilon} \right).$$

Then
$$\frac{\partial \varphi}{\partial x_i}(x) = \frac{1}{\varepsilon} \frac{\partial \phi}{\partial x_i} \left(\frac{x - y}{\varepsilon} \right), \quad i = 1, \dots, n.$$

■

Next we derive a bound so that we may apply the dominated convergence theorem. By the fundamental theorem of calculus

$$\begin{aligned} \varphi(x + he_i) - \varphi(x) &= \int_0^h \frac{\partial}{\partial t}(\varphi(x + te_i))\, dt \\ &= \int_0^h D\varphi(x + te_i) \cdot e_i\, dt, \end{aligned}$$

where $D\varphi = \left(\frac{\partial\varphi}{\partial x_1}, \dots, \frac{\partial\varphi}{\partial x_n}\right)$ is the gradient of φ. This implies that

$$\begin{aligned}
|\varphi(x+he_i)-\varphi(x)| &\leqslant \int_0^{|h|} |D\varphi(x+te_i)\cdot e_i|\,dt \\
&= \frac{1}{\varepsilon}\int_0^{|h|} \left|D\phi\left(\frac{x+te_i-y}{\varepsilon}\right)\cdot e_i\right| dt \\
&\leqslant \frac{1}{\varepsilon}\int_0^{|h|} \left|D\phi\left(\frac{x+te_i-y}{\varepsilon}\right)\right| dt \\
&\leqslant \frac{|h|}{\varepsilon}\|D\phi\|_\infty, \quad i=1,\dots,n.
\end{aligned}$$

Let

$$K = \left\{ y \in \mathbb{R}^n : \frac{x-y}{\varepsilon} \in \operatorname{supp}\phi \text{ or } \frac{x+he_i-y}{\varepsilon} \in \operatorname{supp}\phi,\ 0<|h|<1 \right\}.$$

Since $\operatorname{supp}\phi$ is compact, we see that K is a bounded set. By the estimate above, we have

$$\left|\frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right)-\phi\left(\frac{x-y}{\varepsilon}\right)\right] f(y)\right| \leqslant \frac{1}{\varepsilon}\|D\phi\|_\infty |f(y)|, \quad i=1,\dots,n,$$

for almost every $y \in K$ and we note that $\frac{1}{\varepsilon}\|D\phi\|_\infty |f| \in L^1(K)$. The dominated convergence theorem implies that

$$\begin{aligned}
\frac{\partial(f*\phi_\varepsilon)}{\partial x_i}(x) &= \lim_{h\to 0}\frac{(f*\phi_\varepsilon)(x+he_i)-(f*\phi_\varepsilon)(x)}{h} \\
&= \lim_{h\to 0}\frac{1}{\varepsilon^n}\int_K \frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right)-\phi\left(\frac{x-y}{\varepsilon}\right)\right] f(y)\,dy \\
&= \frac{1}{\varepsilon^n}\int_K \lim_{h\to 0}\frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right)-\phi\left(\frac{x-y}{\varepsilon}\right)\right] f(y)\,dy \\
&= \frac{1}{\varepsilon^n}\int_K \frac{1}{\varepsilon}\frac{\partial\phi}{\partial x_i}\left(\frac{x-y}{\varepsilon}\right) f(y)\,dy \\
&= \int_K \frac{\partial\phi_\varepsilon}{\partial x_i}(x-y)f(y)\,dy \\
&= \left(\frac{\partial\phi_\varepsilon}{\partial x_i} * f\right)(x), \quad i=1,\dots,n.
\end{aligned}$$

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 for $1 \leqslant p < \infty$. This result does not hold true for $p=\infty$. This is simply because the uniform limit of continuous functions is itself continuous.

Remark 3.23. 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)

By Theorem 1.57 we know that $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$ and by Remark 3.22 we know that $C^\infty(\mathbb{R}^n) \cap L^p(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$. Next we give an even stronger result.

Theorem 3.24. $C_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leqslant p < \infty$.

THE MORAL: Not only compactly supported continuous functions, but also compactly supported smooth functions are dense in L^p for $1 \leqslant p < \infty$.

Proof. Assume $f \in L^p(\mathbb{R}^n)$ and let $\eta > 0$. Theorem 1.57 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 $\|f - g\|_{L^p(\mathbb{R}^n)} < \frac{\eta}{2}$. Let ϕ_ε be the standard mollifier in Example 3.20. Theorem 3.21 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

$$K_\varepsilon = \{x \in \mathbb{R}^n : \operatorname{dist}(x, K) \leqslant \varepsilon\}$$

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 3.18 there exists $\varepsilon' > 0$ such that

$$\|g - (g * \phi_\varepsilon)\|_p < \frac{\eta}{2} \quad \text{whenever} \quad 0 < \varepsilon < \varepsilon'.$$

Thus

$$\|f - (g * \phi_\varepsilon)\|_p \leqslant \|f - g\|_p + \|g - (g * \phi_\varepsilon)\|_p < \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

□

3.6 Smoothing in an open subset

Next we discuss smoothing in an open subset of $\mathbb{R}^n$. The convolution smoothing techniques apply also in this case with some minor modifications. For an open subset U of $\mathbb{R}^n$ with $\mathbb{R}^n \setminus U \neq \emptyset$, we consider

$$U_\varepsilon = \{x \in U : \operatorname{dist}(x, \partial U) > \varepsilon\}, \quad \varepsilon > 0.$$

Observe that $U = \bigcup_{\varepsilon > 0} U_\varepsilon$. Let $f \in L^1_{\text{loc}}(U)$. The convolution mollification is $f_\varepsilon : U_\varepsilon \to [-\infty, \infty]$,

$$f_\varepsilon(x) = (f * \phi_\varepsilon)(x) = \int_U f(y)\phi_\varepsilon(x - y)\, dy,$$

where $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi \geqslant 0$, $\operatorname{supp}\phi \subset \overline{B(0,1)}$ and $\int_{\mathbb{R}^n} \phi\, dx = 1$. Here ϕ_ε, $\varepsilon > 0$, is an approximation of the identity as in Definition 3.8. For example, we may consider the standard mollifier as in Example 3.20.

THE MORAL: Since the convolution is a weighted integral average of f over the ball $B(x,\varepsilon)$ for every x, instead of U it is well defined only in U_ε. Sometimes we may consider the zero extension of f to $\mathbb{R}^n \setminus U$. If $U = \mathbb{R}^n$, we do not have this difficulty.

Remark 3.25. Let $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi \geqslant 0$, $\operatorname{supp}\phi \subset \overline{B(0,1)}$ and $\int_{\mathbb{R}^n} \phi\,dx = 1$.

(1) For every $x \in U_\varepsilon$, we have

$$f_\varepsilon(x) = \int_U f(y)\phi_\varepsilon(x-y)\,dy = \int_{B(x,\varepsilon)} f(y)\phi_\varepsilon(x-y)\,dy.$$

(2) By a change of variables $z = x - y$ we have

$$\int_U f(y)\phi_\varepsilon(x-y)\,dy = \int_U f(x-z)\phi_\varepsilon(z)\,dz$$

(3) For every $x \in U_\varepsilon$, we have

$$|f_\varepsilon(x)| \leqslant \left| \int_{B(x,\varepsilon)} f(y)\phi_\varepsilon(x-y)\,dy \right| \leqslant \|\phi_\varepsilon\|_\infty \int_{B(x,\varepsilon)} |f(y)|\,dy < \infty.$$

(4) If $f \in C_0(U)$, then $f_\varepsilon \in C_0(U_\varepsilon)$, whenever $0 < \varepsilon < \varepsilon_0 = \frac{1}{2}\operatorname{dist}(\operatorname{supp} f, \partial U)$.

Reason. If $x \in U_\varepsilon$ such that $\operatorname{dist}(x, \operatorname{supp} f) > \varepsilon_0$ (in particular, for every $x \in U_\varepsilon \setminus U_{\varepsilon_0}$) then $B(x,\varepsilon) \cap \operatorname{supp} f = \emptyset$, which implies that $f_\varepsilon(x) = 0$. ■

We collect properties of the convolution approximation below. The main difference to Theorem 3.14 and Theorem 3.18 is that here we consider compactly supported smooth approximations of the unity instead of more general integrable functions. This simplifies some of the arguments. We denote $U' \Subset U$, if U, U' are open subset of $\mathbb{R}^n$ and $\overline{U'}$ is a compact subset of U. In particular, it follows that $\operatorname{dist}(\overline{U'}, \partial U) > 0$, if $U \neq \mathbb{R}^n$.

Lemma 3.26. Let $U \subset \mathbb{R}^n$ be an open set and assume that $f \in L^p(U)$, $1 \leqslant p < \infty$. Let $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi \geqslant 0$, $\operatorname{supp}\phi \subset \overline{B(0,1)}$ and $\int_{\mathbb{R}^n} \phi\,dx = 1$.

(1) $f_\varepsilon \in C^\infty(U_\varepsilon)$, $\varepsilon > 0$.
(2) $f_\varepsilon \to f$ almost everywhere in U as $\varepsilon \to 0$.
(3) If $f \in C(U)$, then $f_\varepsilon \to f$ uniformly in every $U' \Subset U$ as $\varepsilon \to 0$.
(4) If $f \in L^p_{\mathrm{loc}}(U)$, $1 \leqslant p < \infty$, then $f_\varepsilon \to f$ in $L^p(U')$ for every $U' \Subset U$ as $\varepsilon \to 0$.

Proof. (1) The proof is very similar to the proof of Theorem 3.21. Let $x \in U_\varepsilon$ and $e_i = (0, \dots, 1, \dots, 0)$, $i = 1, \dots, n$. Let $h_0 > 0$ such that $B(x, h_0) \subset U_\varepsilon$ and let $h \in \mathbb{R}$ with $|h| < h_0$. Then

$$\frac{f_\varepsilon(x+he_i) - f_\varepsilon(x)}{h} = \frac{1}{\varepsilon^n} \int_{B(x+he_i,\varepsilon)\cup B(x,\varepsilon)} \frac{1}{h}\left[\phi\left(\frac{x+he_i-y}{\varepsilon}\right) - \phi\left(\frac{x-y}{\varepsilon}\right)\right] f(y)\,dy.$$

Let $U' = B(x, h_0 + \varepsilon)$. Then $U' \Subset U$ and $B(x + he_i, \varepsilon) \cup B(x, \varepsilon) \subset U'$. As in the proof of Theorem 3.21 we may apply the dominated convergence theorem to obtain

$$\begin{aligned}
\frac{\partial f_\varepsilon}{\partial x_i}(x) &= \lim_{h\to 0} \frac{f_\varepsilon(x + he_i) - f_\varepsilon(x)}{h} \\
&= \lim_{h\to 0} \frac{1}{\varepsilon^n} \int_{U'} \frac{1}{h} \left[\phi\left(\frac{x + he_i - y}{\varepsilon}\right) - \phi\left(\frac{x-y}{\varepsilon}\right) \right] f(y)\,dy \\
&= \frac{1}{\varepsilon^n} \int_{U'} \frac{1}{\varepsilon} \frac{\partial \phi}{\partial x_i} \left(\frac{x-y}{\varepsilon}\right) f(y)\,dy \\
&= \int_{U'} \frac{\partial \phi_\varepsilon}{\partial x_i}(x - y) f(y)\,dy \\
&= \left(\frac{\partial \phi_\varepsilon}{\partial x_i} * f\right)(x), \quad i = 1, \dots, n.
\end{aligned}$$

A similar argument shows that $D^\alpha f_\varepsilon$ exists and $D^\alpha f_\varepsilon = D^\alpha \phi_\varepsilon * f$ in U_ε for every multi-index α.

(2) Recall that $\int_{B(x,\varepsilon)} \phi_\varepsilon(x - y)\,dy = 1$. Therefore we have

$$\begin{aligned}
|f_\varepsilon(x) - f(x)| &= \left| \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y) f(y)\,dy - f(x) \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)\,dy \right| \\
&= \left| \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)(f(y) - f(x))\,dy \right| \\
&\leqslant \frac{1}{\varepsilon^n} \int_{B(x,\varepsilon)} \phi\left(\frac{x-y}{\varepsilon}\right) |f(y) - f(x)|\,dy \\
&\leqslant \Omega_n \|\phi\|_{L^\infty(\mathbb{R}^n)} \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} |f(y) - f(x)|\,dy \xrightarrow{\varepsilon \to 0} 0
\end{aligned}$$

for almost every $x \in U$. Here $\Omega_n = |B(0,1)|$ and the last convergence follows from the Lebesgue differentiation theorem (Theorem 2.24).

(3) Let $U' \Subset U'' \Subset U$, $0 < \varepsilon < \operatorname{dist}(U', \partial U'')$, and $x \in U'$. Because $\overline{U''}$ is compact and $f \in C(U)$, f is uniformly continuous in U'', that is, for every $\varepsilon' > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon'$ for every $x, y \in U''$ with $|x - y| < \delta$. By combining this with an estimate from the proof of claim (2), we conclude that

$$|f_\varepsilon(x) - f(x)| \leqslant \Omega_n \|\phi\|_{L^\infty(\mathbb{R}^n)} \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} |f(y) - f(x)|\,dy < \Omega_n \|\phi\|_{L^\infty(\mathbb{R}^n)} \varepsilon'$$

for every $x \in U'$ if $\varepsilon < \delta$.

(4) Let $U' \Subset U'' \Subset U$.

CLAIM:

$$\int_{U'} |f_\varepsilon|^p\,dx \leqslant \int_{U''} |f|^p\,dx$$

whenever $0 < \varepsilon < \operatorname{dist}(U', \partial U'')$ and $0 < \varepsilon < \operatorname{dist}(U'', \partial U)$.

Reason. Let $x \in U'$. By Hölder's inequality, we have

$$\begin{aligned}
|f_\varepsilon(x)| &= \left| \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y) f(y)\,dy \right| \\
&\leqslant \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)^{1-\frac{1}{p}} \phi_\varepsilon(x-y)^{\frac{1}{p}} |f(y)|\,dy \\
&\leqslant \left(\int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)\,dy \right)^{\frac{1}{p'}} \left(\int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)|f(y)|^p\,dy \right)^{\frac{1}{p}} \\
&= \left(\int_{B(x,\varepsilon)} \phi_\varepsilon(y)\,dy \right)^{\frac{1}{p'}} \left(\int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)|f(y)|^p\,dy \right)^{\frac{1}{p}} \\
&= \left(\int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)|f(y)|^p\,dy \right)^{\frac{1}{p}}.
\end{aligned}$$

Here we used the fact that $\int_{\mathbb{R}^n} \phi_\varepsilon\,dx = \int_{\mathbb{R}^n} \phi\,dx = 1$. By raising the previous estimate to power p and by integrating over U', we obtain

$$\begin{aligned}
\int_{U'} |f_\varepsilon(x)|^p\,dx &\leqslant \int_{U'} \int_{B(x,\varepsilon)} \phi_\varepsilon(x-y)|f(y)|^p\,dy\,dx \\
&= \int_{U''} \int_{U'} \phi_\varepsilon(x-y)|f(y)|^p\,dx\,dy \\
&= \int_{U''} |f(y)|^p \int_{U'} \phi_\varepsilon(x-y)\,dx\,dy \\
&= \int_{U''} |f(y)|^p\,dy.
\end{aligned}$$

Here we used Fubini's theorem and once more the fact that $\int_{\mathbb{R}^n} \phi_\varepsilon\,dx = \int_{\mathbb{R}^n} \phi\,dx = 1$. ■

Since $C(U'')$ is dense in $L^p(U'')$. Therefore for every $\varepsilon' > 0$ there exists $g \in C(U'')$ such that

$$\left(\int_{U''} |f-g|^p\,dx \right)^{\frac{1}{p}} \leqslant \frac{\varepsilon'}{3}.$$

By (3), we have $g_\varepsilon \to g$ uniformly in U' as $\varepsilon \to 0$. Thus

$$\left(\int_{U''} |g_\varepsilon - g|^p\,dx \right)^{\frac{1}{p}} \leqslant \sup_{U'} |g_\varepsilon - g| \, |U'|^{\frac{1}{p}} < \frac{\varepsilon'}{3},$$

when $\varepsilon > 0$ is small enough. Now we use Minkowski's inequality and the previous claim to conclude that

$$\begin{aligned}
\left(\int_{U'} |f_\varepsilon - f|^p\,dx \right)^{\frac{1}{p}} &\leqslant \left(\int_{U'} |f_\varepsilon - g_\varepsilon|^p\,dx \right)^{\frac{1}{p}} \\
&\quad + \left(\int_{U'} |g_\varepsilon - g|^p\,dx \right)^{\frac{1}{p}} + \left(\int_{U'} |g - f|^p\,dx \right)^{\frac{1}{p}} \\
&\leqslant 2 \left(\int_{U''} |g - f|^p\,dx \right)^{\frac{1}{p}} + \left(\int_{U'} |g_\varepsilon - g|^p\,dx \right)^{\frac{1}{p}} \\
&\leqslant 2\frac{\varepsilon'}{3} + \frac{\varepsilon'}{3} = \varepsilon'.
\end{aligned}$$

Thus $f_\varepsilon \to f$ in $L^p(U')$ as $\varepsilon \to 0$. □

Next we discuss the density of continuous functions in L^p for open subsets, compare to Remark 3.22. We apply a version of partition of unity in the argument.. We return to this topic in the next section.

Theorem 3.27. Let U be an open subset of $\mathbb{R}^n$. Then $C^\infty(U) \cap L^p(U)$ is dense in $L^p(U)$ for $1 \leqslant p < \infty$.

Proof. Let $U_0 = \emptyset$ and

$$U_i = \left\{x \in U : \operatorname{dist}(x, \partial U) > \tfrac{1}{i}\right\} \cap B(0,i), \quad i = 1,2,\dots.$$

Then $U = \bigcup_{i=1}^\infty U_i$, U_i s open and $\overline{U_i}$ is a compact subset of U_{i+1} for every $i = 1,2,\dots$.

C L A I M : There exists $\varphi_i \in C_0^\infty(U_{i+2} \setminus \overline{U_{i-1}})$, $0 \leqslant \varphi_i \leqslant 1$, $i = 1,2,\dots$ such that

$$\sum_{i=1}^\infty \varphi_i = 1 \quad \text{in} \quad U.$$

Reason. By applying the distance function, we may construct a function $g_i \in C_0^\infty(U_{i+2} \setminus \overline{U_{i-1}})$ with $0 \leqslant g_i \leqslant 1$ and $g_i = 1$ in $\overline{U_{i+1}} \setminus U_i$ for every $i = 1,2,\dots$. Let $\varphi_i : U \to \mathbb{R}$,

$$\varphi_i(x) = \frac{g_i(x)}{\sum_{j=1}^\infty g_j(x)}, \quad i = 1,2,\dots.$$

■

Let $f \in L^p(U)$ with $1 \leqslant p < \infty$ and let ϕ_ε be an approximation of the identity as in Definition 3.8. 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

$$\|\phi_{\varepsilon_i} * (\varphi_i f) - \varphi_i f\|_{L^p(U)} < \frac{\varepsilon}{2^i}, \quad i = 1,2,\dots.$$

Let

$$g = \sum_{i=1}^\infty \phi_{\varepsilon_i} * (\varphi_i f).$$

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, we have

$$\begin{aligned}\|f - g\|_{L^p(U)} &= \left\| \sum_{i=1}^\infty \phi_{\varepsilon_i} * (\varphi_i f) - \sum_{i=1}^\infty \varphi_i f \right\|_{L^p(U)} \\ &\leqslant \sum_{i=1}^\infty \left\| \phi_{\varepsilon_i} * (\varphi_i f) - \varphi_i f \right\|_{L^p(U)} \leqslant \sum_{i=1}^\infty \frac{\varepsilon}{2^i} = \varepsilon.\end{aligned}$$

This shows that $C^\infty(U)$ is dense in $L^p(U)$ for $1 \leqslant p < \infty$ for an arbitrary open subset U of $\mathbb{R}^n$. □

Next we discuss a version of Theorem 3.24 for an open subset of $\mathbb{R}^n$. In the proof of the previous theorem, we worked inside the open set throughout. In the proof of the next result we apply zero extension to the complement. This proof can also be arranged so that we work inside the open set throughout (exercise).

Theorem 3.28. Let U be an open subset of $\mathbb{R}^n$. Then $C_0^\infty(U)$ is dense in $L^p(U)$ for $1 \leqslant p < \infty$.

Proof. Let $f \in L^p(U)$ and extend as zero to $\mathbb{R}^n \setminus U$. Let

$$U_i = \left\{x \in U : \operatorname{dist}(x, \mathbb{R}^n \setminus U) > \tfrac{2}{i}\right\} \cap B(0,i), \quad i = 1,2,\dots.$$

Then $U_i \subset U_{i+1}$, $\overline{U_i}$ is a compact subset of U_{i+1} for every $i = 1,2,\dots$ and $U = \bigcup_{i=1}^\infty U_i$. Let

$$g_i = f\chi_{U_i} \quad \text{and} \quad f_i = \phi_{\frac{1}{i}} * g_i, \quad i = 1,2,\dots,$$

where $\phi \in C_0^\infty(\mathbb{R}^n)$ with $\phi \geqslant 0$, $\operatorname{supp}\phi \subset \overline{B(0,1)}$ and $\int_{\mathbb{R}^n} \phi\,dx = 1$. Here $\phi_{\frac{1}{i}}$ is an approximation of the identity as in Definition 3.8 with $\varepsilon = \frac{1}{i}$. For example, we may consider the standard mollifier in Example 3.20. Since $\operatorname{supp}\phi_{\frac{1}{i}} \subset \overline{B(0,\frac{1}{i})}$ it follows that

$$\operatorname{supp} f_i \subset \left\{x \in \mathbb{R}^n : \operatorname{dist}(x, U_i) \leqslant \tfrac{1}{i}\right\} \subset U$$

for every $i = 1,2,\dots$. Note that $\left\{x \in \mathbb{R}^n : \operatorname{dist}(x, U_i) \leqslant \frac{1}{i}\right\}$ is a closed and bounded set and thus a compact subset of U. Consequently $\operatorname{supp} f_i$ is a compact subset of U for every $i = 1,2,\dots$. It follows from Lemma 3.26 (1) that $f_i \in C_0^\infty(U)$ for every $i = 1,2,\dots$.

By Minkowski's inequality and Young's theorem (Theorem 3.4), we have

$$\begin{aligned}
\|f_i - f\|_{L^p(U)} &= \|f_i - f\|_{L^p(\mathbb{R}^n)} = \|\phi_{\frac{1}{i}} g_i - f\|_{L^p(\mathbb{R}^n)} \\
&\leqslant \|\phi_{\frac{1}{i}} * g_i - \phi_{\frac{1}{i}} * f\|_{L^p(\mathbb{R}^n)} + \|\phi_{\frac{1}{i}} * f - f\|_{L^p(\mathbb{R}^n)} \\
&\leqslant \|\phi_{\frac{1}{i}}\|_1 \|g_i - f\|_{L^p(\mathbb{R}^n)} + \|\phi_{\frac{1}{i}} * f - f\|_{L^p(\mathbb{R}^n)} \\
&= \|f\chi_{U_i} - f\|_{L^p(\mathbb{R}^n)} + \|\phi_{\frac{1}{i}} * f - f\|_{L^p(\mathbb{R}^n)}
\end{aligned}$$

for every $i = 1,2,\dots$. Here we also applied the fact that $\|\phi_{\frac{1}{i}}\|_1 = 1$ for every $i = 1,2,\dots$. Since $|f\chi_{U_i}| \leqslant |f|$ for every $i = 1,2,\dots$, by the dominated convergence theorem (Theorem 1.37), we have

$$\|f\chi_{U_i} - f\|_{L^p(\mathbb{R}^n)} \xrightarrow{i\to\infty} 0.$$

On the other hand, Theorem 3.18 implies that

$$\|\phi_{\frac{1}{i}} * f - f\|_{L^p(\mathbb{R}^n)} \xrightarrow{i\to\infty} 0.$$

It follows that

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

This completes the proof. □

3.7 Partition of unity

In this section we briefly discuss partition of unity which is a useful tool to localize problems in analysis. We already applied a partition of unity in the proof of Theorem 3.27.

Theorem 3.29. Let $U \subset \mathbb{R}^n$ be an open set and let $\{U_\alpha\}_{\alpha\in I}$ be an open covering of U. There exist functions $\varphi_i \in C_0^\infty(U)$, $i = 1, 2, \dots$, such that

(1) $\sum_{i=1}^{k} \varphi_i(x) = 1$ for every $x \in U$,

(2) $\operatorname{supp}\varphi_i$ is a subset of U_α for some $\alpha \in I$ and

(3) for every compact set $K \subset U$, there exists an integer k and an open set U' with $K \subset U' \subset U$ such that $\sum_{i=1}^{k} \varphi_i(x) = 1$ for every $x \in U'$.

T E R M I N O L O G Y : The collection of functions φ_i is called a partition of unity related to the covering $\{U_\alpha\}_{\alpha\in I}$. Observe that I is an arbitrary index set which is not necessarily countable.

T H E M O R A L : Partition of unity is a very useful tool to localize functions, since

$$f(x) = f(x)\sum_{i=1}^{k} \varphi_i(x) = \sum_{i=1}^{k} f(x)\varphi_i(x) \quad \text{for every} \quad x \in U'.$$

Thus a function can be represented as a sum of compactly supported functions.

Proof. (1) Let S be a countable dense subset of U. For example, we may take $S = \{x = (x_1, \dots, x_n) \in U : x_i \in \mathbb{Q}, i = 1, \dots, n\}$. Consider a collection $\mathscr{F}$ of countably many closed balls

$$\mathscr{F} = \left\{\overline{B(x_i, r_i)} : 0 < r_i < 1, r_i \in \mathbb{Q}, x_i \in S, \overline{B(x_i, r_i)} \subset U_\alpha \cap U \text{ for some } \alpha \in I\right\}.$$

Since$\{U_\alpha\}_{\alpha\in I}$ is an open covering of U, by the density of S in U and the density of the rational numbers in the real line, we have

$$U = \bigcup_{i=1}^{\infty} B\left(x_i, \tfrac{r_i}{2}\right),$$

that is, the set U is a countable union of the respective open balls in $\mathscr{F}$.

(2) Let

$$g_i = \phi_{\frac{r_i}{4}} * \chi_{B(x_i, \frac{3}{4} r_i)}, \quad i = 1, 2, \dots,$$

where ϕ is the standard mollifier in Example 3.20. Here $\phi_{\frac{r_i}{4}}$ is an approximation of the identity as in Definition 3.8 with $\varepsilon = \frac{r_i}{4}$. Theorem 3.21 implies that $g_i \in C^\infty(\mathbb{R}^n)$,

$i = 1,2,\dots$. We note that if $x \in B(x_i, \frac{r_i}{2})$, then $B(x, \frac{r_i}{4}) \subset B(x_i, \frac{3}{4}r_i)$. This implies that

$$\begin{aligned} g_i(x) = \phi_{\frac{r_i}{4}} * \chi_{B(x_i,\frac{3}{4}r_i)}(x) &= \int_{\mathbb{R}^n} \phi_{\frac{r_i}{4}}(x-y)\chi_{B(x_i,\frac{3}{4}r_i)}(y)\,dy \\ &= \int_{B(x,\frac{r_i}{4})} \phi_{\frac{r_i}{4}}(x-y)\chi_{B(x_i,\frac{3}{4}r_i)}(y)\,dy \\ &= \int_{B(x,\frac{r_i}{4})} \phi_{\frac{r_i}{4}}(x-y)\,dy = 1 \end{aligned}$$

for every $x \in B(x_i, \frac{r_i}{2})$. Here we also used the properties

$$\operatorname{supp}\phi_{\frac{r_i}{4}} = \overline{B(0,\frac{r_i}{4})} \quad \text{and} \quad \|\phi_{\frac{r_i}{4}}\|_1 = 1, \quad i = 1,2,\dots.$$

Since $0 \leqslant \chi_{B(x_i,\frac{3}{4}r_i)} \leqslant 1$ and $\phi \geqslant 0$, a similar argument as above shows that $0 \leqslant g_i \leqslant 1$, $i = 1,2,\dots$. Moreover, if $x \notin B(x_i, r_i)$, then $B(x, \frac{r_i}{4}) \cap B(x_i, \frac{3}{4}r_i) = \emptyset$. This implies that

$$g_i(x) = \int_{B(x,\frac{r_i}{4})} \phi_{\frac{r_i}{4}}(x-y)\chi_{B(x_i,\frac{3}{4}r_i)}(y)\,dy = 0$$

for every $x \notin B(x_i, r_i)$. It follows that $g_i \in C_0^\infty(\mathbb{R}^n)$ and $\operatorname{supp} g_i \subset \overline{B(x_i,r_i)}$. By the definition of $\mathscr{F}$, we conclude that $\operatorname{supp} g_i \subset \overline{B(x_i,r_i)} \subset U_\alpha \cap U$ for some $\alpha \in I$.

(3) Let

$$\begin{aligned} \varphi_1 &= g_1, \\ \varphi_2 &= (1-g_1)g_2, \\ &\vdots \\ \varphi_k &= (1-g_1)\dots(1-g_{k-1})g_k. \end{aligned}$$

Since $0 \leqslant g_i \leqslant 1$ and $\operatorname{supp} g_i \subset \overline{B(x_i,r_i)}$, $i = 1,2,\dots$, we have $0 \leqslant \varphi_i \leqslant 1$ and $\operatorname{supp}\varphi_i \subset \overline{B(x_i,r_i)}$, $i = 1,2,\dots$.

(4) We show by induction that

$$\sum_{i=1}^{k} \varphi_i = 1 - (1-g_1)\dots(1-g_k), \quad k = 1,2,\dots.$$

This is true for $k = 1$, since $\varphi_1 = g_1$. Assume that the formula above holds true for some k. Then

$$\begin{aligned} \sum_{i=1}^{k+1} \varphi_i &= 1 - (1-g_1)\dots(1-g_k) + \varphi_{k+1} \\ &= 1 - (1-g_1)\dots(1-g_k) + (1-g_1)\dots(1-g_k)g_{k+1} \\ &= 1 - (1-g_1)\dots(1-g_{k+1}). \end{aligned}$$

(5) Since $g_i = 1$ in $B(x_i, \frac{r_i}{2})$, $i = 1,2,\dots$, we have

$$\sum_{i=1}^{j} \varphi_i(x) = 1 - (1-g_1(x))\dots(1-g_j(x)) = 1$$

for every $x \in \bigcup_{i=1}^{k} B(x_i, \frac{r_i}{2})$ and every $j \geqslant k$. Since $U = \bigcup_{i=1}^{\infty} B(x_i, \frac{r_i}{2})$, for every $x \in U$ there exists k such that $x \in \bigcup_{i=1}^{k} B(x_i, \frac{r_i}{2})$. It follows that $\sum_{i=1}^{j} \varphi_i(x) = 1$ for every $j \geqslant k$ and thus

$$\sum_{i=1}^{\infty} \varphi_i(x) = \lim_{j\to\infty} \sum_{i=1}^{j} \varphi_i(x) = 1.$$

Finally, let K be a compact subset of U. Since $\{B(x_i, \frac{r_i}{2})\}$ is an open covering of K, there exists a finite subcovering such that $K \subset \bigcup_{i=1}^{k} B(x_i, \frac{r_i}{2})$. Let $U' = \bigcup_{i=1}^{k} B(x_i, \frac{r_i}{2})$. As a finite union of open balls the set U' is open. Since $B(x_i, \frac{r_i}{2}) \subset U$, $i = 1, \dots, k$, we have $K \subset U' \subset U$. Moreover, we have

$$\sum_{i=1}^{k} \varphi_i(x) = 1 - (1 - g_1(x)) \dots (1 - g_j(x)) = 1$$

for every $x \in \bigcup_{i=1}^{k} B(x_i, \frac{r_i}{2}) = U$. □

Remark 3.30. The smoothing process with convolutions is applied to construct a smooth partition of unity. If we are only interested in having a partition of unity by compactly supported continuous functions, that is, $\varphi_i \in C(U)$, $i = 1, 2, \dots$, we can construct the required cutoff functions by applying the distance function as in (1.58).

3.8 The Poisson kernel

We consider an example from the theory of partial differential equations. Assume that $f \in L^p(\mathbb{R}^n)$, $1 \leqslant p < \infty$. Let $P : \mathbb{R}^n \to \mathbb{R}$,

$$P(x) = c(n)(1 + |x|^2)^{-\frac{n+1}{2}}, \quad c(n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{\frac{n+1}{2}}},$$

be the Poisson kernel, where the constant $c(n)$ is chosen such that

$$\int_{\mathbb{R}^n} P(x)\,dx = 1.$$

Then

$$P_\varepsilon(x) = \frac{1}{\varepsilon^n} P\left(\frac{x}{\varepsilon}\right) = c(n)\varepsilon(|x|^2 + \varepsilon^2)^{-\frac{n+1}{2}}, \quad \varepsilon > 0,$$

is an approximation of the identity and we may apply the theory developed above. By Young's theorem (Theorem 3.4), 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 3.14 implies that

$$\lim_{\varepsilon\to 0}(f * P_\varepsilon)(x) = f(x) \quad \text{for almost every} \quad x \in \mathbb{R}^n.$$

It can be shown that the function $x \mapsto u(x,\varepsilon) = (f * P_\varepsilon)(x)$ belongs to $C^\infty(\mathbb{R}^n)$ for every $\varepsilon > 0$ (exercise). Observe that we cannot directly apply Theorem 3.21, since the Poisson kernel is not compactly supported. Moreover, 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 3.18 shows that $u(x,\varepsilon) \to f(x)$ in $L^p(\mathbb{R}^n)$ as $\varepsilon \to 0$. Note also, that by Theorem 3.13, 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 3.31. 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 later.

We discuss the covering lemma, see Theorem 2.15. For an arbitrary Radon measure μ on $\mathbb{R}^n$, there is no uniform way to control $\mu(B(x,2r))$ in terms of $\mu(B(x,r))$. The measure μ is called doubling, if there is a constant c such that

$$\mu(B(x,2r)) \leqslant c\mu(B(x,r)) \quad \text{for every} \quad x \in \mathbb{R}^n, r > 0.$$

The factor two does not play any particular role in the doubling condition and it can be replaced by any factor that is strictly greater than one. For example, we have

$$\begin{aligned}\mu(B(x,5r)) &\leqslant c\mu(B(x,\tfrac{5}{2}r)) \leqslant c^2\mu(B(x,\tfrac{5}{4}r))\\ &\leqslant c^3\mu(B(x,\tfrac{5}{8}r)) \leqslant c^3\mu(B(x,r)) \quad \text{for every} \quad x\in\mathbb{R}^n, r>0.\end{aligned}$$

Let A be a bounded subset of $\mathbb{R}^n$ and assume 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 covering lemma, see Theorem 2.15, 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, Theorem 2.15 is not useful for a general Radon measure. 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 $P = P(n)$ and $Q = Q(n)$ 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 exists 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 exist 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

$$A \subset \bigcup_{k=1}^{Q}\bigcup_{\mathscr{F}_k} B(x_i,r_i).$$

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. The main advantage compared to the covering lemma, see Theorem 2.15, is that we do not have to enlarge the covering balls.

Let A is a bounded subset of $\mathbb{R}^n$ and assume 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 A and by the bounded overlap property, we have

$$\sum_{i=1}^{\infty} \chi_{B(x_i,r_i)}(x) \leqslant P \chi_{\bigcup_{i=1}^{\infty} B(x_i,r_i)}(x)$$

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

$$\begin{aligned}
\mu(A) &\leqslant \mu\Big(\bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) \leqslant \sum_{i=1}^{\infty} \mu(B(x_i,r_i)) = \sum_{i=1}^{\infty} \int_{\mathbb{R}^n} \chi_{B(x_i,r_i)}(x)\,d\mu(x) \\
&= \int_{\mathbb{R}^n} \sum_{i=1}^{\infty} \chi_{B(x_i,r_i)}(x)\,d\mu(x) \leqslant P \int_{\mathbb{R}^n} \chi_{\bigcup_{i=1}^{\infty} B(x_i,r_i)}(x)\,d\mu(x) \\
&= P\mu\Big(\bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) \leqslant P\mu\Big(\bigcup_{x\in A} B(x,r_x)\Big).
\end{aligned}$$

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 $\{B_i\}$ 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 as in the covering lemma, see Theorem 2.15.

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

$$\begin{aligned}\left|\frac{x}{|x|}-\frac{y}{|y|}\right| \geqslant 1 &\iff 2-2\frac{\langle x,y\rangle}{|x||y|}\geqslant 1 \\ &\iff \frac{\langle x,y\rangle}{|x||y|}\leqslant\frac{1}{2} \\ &\iff \cos\angle(x,y)\leqslant\frac{1}{2} \\ &\iff \angle(x,y)\geqslant 60°\end{aligned}$$

■

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 $\angle(x,y)\geqslant 60°$ (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 $k\leqslant N(n)$.

THE MORAL: Condition (1) asserts that the center of every ball belongs only to that ball of which center it is and (2) asserts that all balls intersect at some point. The claim is that there can be only a bounded number of 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

$$\left| \frac{x_i}{|x_i|} - \frac{x_j}{|x_j|} \right| \geqslant 1, \quad j \neq i \tag{4.7}$$

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 (4.7). □

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 $r_x > 2\operatorname{diam} A$, 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 $B\big(x_i, \frac{M_1}{4}\big)$, $i = 1,2,\dots$, are disjoint. Then $B\big(x_i, \frac{M_1}{4}\big) \subset B(x,R)$ with $R = \operatorname{diam}(A) + M$ and $x \in A$. This implies

$$\sum_{i=1}^{k} \left|B\big(x_i, \tfrac{M_1}{4}\big)\right| = \left|\bigcup_{i=1}^{k} B\big(x_i, \tfrac{M_1}{4}\big)\right| \leqslant |B(x,R)|.$$

On the other hand,

$$\sum_{i=1}^{k} \left|B\left(x_i, \tfrac{M_1}{4}\right)\right| = k|B(0,1)|\left(\tfrac{M_1}{4}\right)^n$$

which implies

$$k \leqslant \left(\tfrac{4}{M_1}\right)^n \frac{|B(x,R)|}{|B(0,1)|} < \infty$$

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

$$x_{k_1+1} \in A \setminus \bigcup_{i=1}^{k_1} B(x, r_{x_i}) \quad \text{such that} \quad r_{x_{k_1+1}} \geqslant \tfrac{M_2}{2}$$

and recursevely

$$x_{j+1} \in A \setminus \bigcup_{i=1}^{j} B(x_i, r_{x_i}) \quad \text{such that} \quad r_{x_{j+1}} \geqslant \tfrac{M_2}{2}.$$

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 $M_{i+1} \leqslant \frac{M_i}{2}$,
(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:

$$\tfrac{M_j}{2} \leqslant r_{x_i} \leqslant M_j \leqslant \tfrac{M_{j-1}}{2}, \quad \text{when} \quad i \in I_j, \tag{4.8}$$

$$x_{j+1} \in A \setminus \bigcup_{i=1}^{j} B(x_i, r_{x_i}) \quad \text{and} \tag{4.9}$$

$$x_i \in A \setminus \bigcup_{m\neq k} \bigcup_{j\in I_m} B(x_j, r_{x_j}), \quad \text{when} \quad i \in I_k. \tag{4.10}$$

Reason. The first two properties (4.8) and (4.9) follow from the construction. To prove (4.10), assume that $i \in I_k$, $m \neq k$ and $j \in I_m$. If $m < k$, then by (4.9) we have $x_i \notin B(x_j, r_{x_j})$. If $k < m$, then (4.8) and $m-1 \geqslant k$ imply

$$r_{x_j} \leqslant M_m \leqslant \tfrac{M_{m-1}}{2} \leqslant \tfrac{M_k}{2} \leqslant r_{x_i}.$$

Thus (4.9) implies $x_j \notin B(x_i, r_{x_i})$ and consequently $x_i \notin B(x_j, r_{x_j})$. ∎

Observe that

$$\begin{aligned}(4.9) \Longrightarrow \quad & M_i \leqslant 2^{1-i} M_1,\ i = 1,2,\ldots \\ \Longrightarrow \quad & M_i \to 0,\ i \to \infty \\ \Longrightarrow \quad & r_{x_i} \to 0,\ i \to \infty.\end{aligned}$$

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 $B_1, \ldots, B_s$ are balls in the collection $\{B(x_{m_i}, r_{x_{m_i}})\}_{i=1,\ldots,p}$ with the property that each ball belongs to a different class of indices I_j. Property (4.10) 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

$$\sharp\big\{j : I_j \cap \{m_i : i = 1, \ldots, p\} \neq \emptyset\big\} \leqslant N(n).$$

In other words, the indices m_i can belong to at most $N(n)$ classes of indices I_j.

CLAIM: $\sharp\{I_j \cap \{m_i : i = 1, \ldots, p\}\} \leqslant 16^n$, $j = 1, 2, \ldots$.

Reason. Fix j and denote

$$I_j \cap \{m_i : i = 1, \ldots, p\} = \{l_1, \ldots l_q\}$$

Properties (4.8) and (4.9) imply $B(x_{l_i}, \frac{1}{4} r_{x_{l_i}})$, $i = 1, \ldots q$, 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 $B(x_i, r_{x_i})$, $i = 1, 2, \ldots$, 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 $r_{x_1} \geqslant r_{x_2} \geqslant \ldots$. Denote $B_i = B(x_i, r_{x_i})$, $i = 1, 2, \ldots$. 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

$$\mathscr{B}_1 = \{B_{1,1}, B_{1,2}, \ldots\},$$

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 4.11:

(1) The assumption that A is bounded in Theorem 4.2 can be replaced with the assumption that the radii of the balls in $\mathscr{F}$ are uniformly bounded, that is, $\sup\{r : B(x,r) \in \mathscr{F}\} < \infty$.

(2) Theorem 4.2 applies also for open balls.

(3) Balls in Theorem 4.2 can be replaced, for example, by cubes.

We take another look at the covering theorem. For the Lebesgue measure, the following covering theorem can be proved by applying Theorem 2.15. For a general Radon measure, we apply the Besicovitch covering theorem instead.

Theorem 4.12 (Infinitesimal 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 exist disjoint balls $B(x_i, r_i) \in \mathscr{F}$, $i = 1, 2, \ldots$, such that

$$\mu\Big(A \setminus \bigcup_{i=1}^{\infty} B(x_i, r_{x_i})\Big) = 0.$$

THE MORAL: The main advantage compared to the Besicovitch covering theorem is that the balls that the covering balls are pairwise disjoint under the assumption that there exist arbitrarily small balls centered at every point.

Remark 4.13. The infinitesimal covering theorem implies that every open set can be exhausted by countably many disjoint balls up to a set of measure zero. Observe that this result holds true not only for the Lebesgue measure, but also for a general Radon measure. Recall that in the one-dimensional case every nonempty open set is a union of countably many disjoint open intervals and in the higher dimensional case every nonempty open set is a union of countably many pairwise disjoint half open dyadic cubes.

Let A be a μ-measurable subset of $\mathbb{R}^n$ and assume that for every $x \in A$ there are balls $B(x,r)$ with arbitrary small radii $r > 0$. By the infinitesimal covering theorem, see Theorem 4.12, we have a countable subcollection of pairwise disjoint balls $B(x_i, r_i)$, $i = 1, 2, \ldots$, such that

$$\mu\Big(A \setminus \bigcup_{i=1}^{\infty} B(x_i, r_i)\Big) = 0.$$

Thus

$$\begin{aligned}\mu(A) &\leqslant \mu\Big(A \cap \bigcup_{i=1}^{\infty} B(x_i, r_i)\Big) + \underbrace{\mu\Big(A \setminus \bigcup_{i=1}^{\infty} B(x_i, r_i)\Big)}_{=0} \\ &\leqslant \sum_{i=1}^{\infty} \mu(B(x_i, r_i)) = \mu\Big(\bigcup_{i=1}^{\infty} B(x_i, r_i)\Big).\end{aligned}$$

Proof. We may assume that $\mu(A) > 0$, because otherwise the claim is clear. Assume first that A is bounded. Then there exists a compact set K such that $A \subset K$ and thus $\mu(A) \leqslant \mu(K) < \infty$. Since μ is Borel regular, there exists a Borel set B such that $A \subset B$ and $\mu(B) = \mu(A)$. Thus we may assume that A is a Borel set and, in particular, A is μ-measurable. By the approximation properties of measurable sets for a Radon measure, there exists an open set $G \supset A$ such that

$$\mu(G) \leqslant \Big(1 + \tfrac{1}{4Q}\Big)\mu(A).$$

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 pairwise disjoint and

$$A \subset \bigcup_{k=1}^{Q} \bigcup_{\mathscr{F}_k} B(x_i, r_i) \subset G.$$

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

$$\mu(A) > Q\mu\Big(\bigcup_{\mathscr{F}_k} B(x_i, r_i)\Big) \quad \text{for every} \quad k = 1,\dots,Q,$$

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 implies

$$\mu(A) > \sum_{k=1}^{Q} \mu\Big(\bigcup_{\mathscr{F}_k} B(x_i, r_i)\Big).$$

■

Since

$$\mu(A) \leqslant Q\mu\Big(\bigcup_{\mathscr{F}_k} B(x_i, r_i)\Big) = Q \sum_{\mathscr{F}_k} \mu(B(x_i, r_i)),$$

there exists a finite subcollection $\mathscr{F}_1' \subset \mathscr{F}_k$ such that

$$Q \sum_{\mathscr{F}_1'} \mu(B(x_i, r_i)) \geqslant \frac{\mu(A)}{2}.$$

This implies

$$\mu(A) \leqslant 2Q \sum_{\mathscr{F}_1'} \mu(B(x_i, r_i)) = 2Q\mu\Big(\bigcup_{\mathscr{F}_1'} B(x_i, r_i)\Big).$$

Let

$$A_1 = A \setminus \bigcup_{\mathscr{F}_1'} B(x_i, r_i).$$

Then

$$\begin{aligned}
\mu(A_1) &\leqslant \mu\Big(G \setminus \bigcup_{\mathscr{F}_1'} B(x_i, r_i)\Big) \\
&= \mu(G) - \mu\Big(\bigcup_{\mathscr{F}_1'} B(x_i, r_i)\Big) \leqslant \Big(1 + \tfrac{1}{4Q} - \tfrac{1}{2Q}\Big)\mu(A) \\
&= \Big(1 - \tfrac{1}{4Q}\Big)\mu(A) = \gamma\mu(A), \quad \gamma = 1 - \frac{1}{4Q} < 1.
\end{aligned}$$

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

$$\mathscr{F}' = \Big\{B(x,r) \in \mathscr{F} : B(x,r) \cap \Big(\bigcup_{\mathscr{F}_1'} B(x_i,r_i)\Big) = \emptyset\Big\}.$$

Note that A_1 is a subset of the open set $G \setminus \bigcup_{\mathscr{F}_1'} B(x_i,r_i)$. 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 (exercise). 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

$$A_1 = \{x \in \mathbb{R}^n : |x| < r_1\}, \quad A_k = \{x \in \mathbb{R}^n : r_{k-1} < |x| < r_k\}, \quad k = 2,3,\ldots$$

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 derivative of a measure. Let $f : [a,b] \to [0,\infty]$ be a nonnegative integrable function and $F : [a,b] \to \mathbb{R}$,

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

By Theorem 2.33, we have $F'(x) = f(x)$ for almost every $x \in [a,b]$. Let us write this in another way. Define a measure by letting $\mu(A) = \int_A f(y)\,dy$ for every Lebesgue measurable set $A \subset \mathbb{R}$ and let ν be the one-dimensional Lebesgue measure. 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 4.14. 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 4.15:

(1) Let $A \subset \mathbb{R}^n$ be μ-measurable. By the measure theory, the restriction $\nu = \mu\lfloor A$ 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 that μ is a Radon measures on $\mathbb{R}^n$ and $f \in L^1(\mathbb{R}^n;\mu)$. Let $\nu(A) = \int_A |f|\,d\mu$ 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.

Recall that a function is Borel measurable, if the preimage of every Borel set is a Borel set.

Lemma 4.16. $\overline{D}_\mu \nu$, $\underline{D}_\mu \nu$ and $D_\mu \nu$ are Borel measurable.

THE MORAL: Derivatives of measures are Borel measurable and, in particular, measurable functions with respect to any Radon measure.

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 an open set $G\supset B(x,r)$ such that $\mu(G)<\mu(B(x,r))+\varepsilon$. Observe that $B(x,r)$ denotes the closed ball with center x and radius r. It follows that $B(y,r)\subset G$, if $|x-y|<\frac{1}{2}\operatorname{dist}(B(x,r),\mathbb{R}^n\setminus G)$. Thus

$$\mu(B(y,r))\leqslant\mu(G)<\mu(B(x,r))+\varepsilon,$$

if $|x-y|<\frac{1}{2}\operatorname{dist}(B(x,r),\mathbb{R}^n\setminus G)$. 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 $x\mapsto\nu(B(x,r))$ is upper semicontinuous and consequently the functions are Borel measurable (exercise). ■

CLAIM: $\overline{D}_\mu\nu(x)=\limsup_{\substack{r\to0\\ r\in\mathbb{Q}_+}}\dfrac{\nu(B(x,r))}{\mu(B(x,r))}$.

Reason. Since $B(x,r)$ is a closed ball,

$$B(x,r)=\bigcap_{i=1}^{\infty}B\left(x,r+\tfrac{1}{i}\right)\quad\text{and}\quad\mu(B(x,r+1))<\infty,$$

we have

$$\mu(B(x,r))=\mu\Big(\bigcap_{i=1}^{\infty}B\left(x,r+\tfrac{1}{i}\right)\Big)=\lim_{i\to\infty}\mu\left(B\left(x,r+\tfrac{1}{i}\right)\right).$$

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 4.17. 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 4.18. The set $A \subset \mathbb{R}^n$ does not necessarily have to be measurable, compare to Theorem 4.12.

Proof. (1) If $\mu(A) = \infty$, the claim is clear, so that we may assume $\mu(A) < \infty$. Let $\varepsilon > 0$. There exists an open set $G \supset A$ such that $\mu(G) < \mu(A) + \varepsilon$. Since $\underline{D}_\mu \nu(x) \leqslant t$ for every $x \in A$, there exists 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 infinitesimal covering theorem (Theorem 4.12), there is a countable subcollection of pairwise disjoint balls $B(x_i, r_i) \subset G$, $i = 1, 2, \dots$, such that

$$\nu(B(x,r_i)) \leqslant (t+\varepsilon)\mu(B(x,r_i))$$

for every $i = 1, 2, \dots$ and

$$\nu\Big(A \setminus \bigcup_{i=1}^{\infty} B(x_i, r_i)\Big) = 0.$$

Thus

$$\begin{aligned}
\nu(A) &\leqslant \nu\Big(A \cap \bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) + \underbrace{\nu\Big(A \setminus \bigcup_{i=1}^{\infty} B(x_i,r_i)\Big)}_{=0} \\
&\leqslant \sum_{i=1}^{\infty} \nu(B(x_i,r_i)) \leqslant (t+\varepsilon)\sum_{i=1}^{\infty} \mu(B(x_i,r_i) \\
&\leqslant (t+\varepsilon)\mu\Big(\bigcup_{i=1}^{\infty} B(x_i,r_i)\Big) \quad \text{(the balls are disjoint)} \\
&\leqslant (t+\varepsilon)\mu(G) \leqslant (t+\varepsilon)(\mu(A)+\varepsilon).
\end{aligned}$$

Letting $\varepsilon \to 0$, we have $\nu(A) \leqslant t\mu(A)$.

(2) (Exercise) □

Theorem 4.19. 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 CLAIM: $\overline{D}_\mu \nu = \underline{D}_\mu \nu$ μ-almost everywhere in $\mathbb{R}^n$, that is,

$$\mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) > \underline{D}_\mu \nu(x)\}) = 0.$$

Reason. Let $i, k \in \{1, 2, \dots\}$, $p, q \in \mathbb{Q}$ with $p < q$. Let

$$A_{k,p,q} = \{x \in B(0,k) : \underline{D}_\mu \nu(x) \leqslant p < q \leqslant \overline{D}_\mu \nu(x)\}$$

and

$$A_{k,i} = \{x \in B(0,k) : \overline{D}_\mu \nu(x) \geqslant i\}.$$

Observe that

$$\{x \in B(0,k) : \overline{D}_\mu \nu(x) > \underline{D}_\mu \nu(x)\} = \bigcup_{\substack{0<p<q \\ p,q,\in\mathbb{Q}}} A_{k,p,q}.$$

Since $\overline{D}_\mu \nu(x) \geqslant q$ for every $x \in A_{k,p,q}$, Theorem 4.17 implies that $\nu(A_{k,p,q}) \geqslant q\mu(A_{k,p,q})$. On the other hand, since $\underline{D}_\mu \nu(x) \leqslant p$ for every $x \in A_{k,p,q}$, Theorem 4.17 implies that $\nu(A_{k,p,q}) \leqslant p\mu(A_{k,p,q})$. Thus we have

$$q\mu(A_{k,p,q}) \leqslant \nu(A_{k,p,q}) \leqslant p\mu(A_{k,p,q}),$$

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 CLAIM: $\overline{D}_\mu \nu < \infty$ μ-almost everywhere in $\mathbb{R}^n$ or equivalently

$$\mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) = \infty\}) = 0.$$

Reason. Theorem 4.17 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

$$\mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}) = 0$$

for every $k = 1,2,\dots$. This implies

$$\begin{aligned} \mu(\{x \in \mathbb{R}^n : \overline{D}_\mu \nu(x) = \infty\}) &= \mu\Big(\bigcup_{k=1}^{\infty} \{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}\Big) \\ &\leqslant \sum_{k=1}^{\infty} \mu(\{x \in B(0,k) : \overline{D}_\mu \nu(x) = \infty\}) = 0. \end{aligned}$$ ■

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 let $\nu(A) = \int_A f\,d\mu$, 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 (Theorem 4.23 below) shows that this is the case if ν is absolutely continuous with respect to μ.

Definition 4.20. 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.

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 $\mu_1 \ll \mu_3$.)

Examples 4.21:

(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. In this case it is not reasonable to expect that there exists a μ-measurable function f such $\nu(A) = \int_A f\,d\mu$ for every μ-measurable set A, since

$$\nu(\{0\}) = \int_{\{0\}} f\,d\mu = 0.$$

(2) Let f be a nonnegative μ-measurable function and let $\nu(A) = \int_A f\,d\mu$, where A is μ-measurable. Then $\mu(A) = 0$ implies $\nu(A) = \int_A f\,d\mu = 0$. Thus $\nu \ll \mu$.

Remark 4.22. 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 $\mu(A) < \delta$.

Reason. $\boxed{\Longleftarrow}$ Assume that the ε, δ-version of absolute continuity holds. Let $A \subset \mathbb{R}^n$ with $\mu(A) = 0$. Then $\nu(A) < \varepsilon$ for every $\varepsilon > 0$ and thus $\nu(A) = 0$. This implies that $\nu \ll \mu$.

$\boxed{\Longrightarrow}$ For a contradiction, assume that $\nu \ll \mu$ and that the ε, δ-version of absolute continuity fails. Then there exist $\varepsilon > 0$ and μ-measurable sets $A_i \subset \mathbb{R}^n$, $i = 1, 2, \dots$, such that $\mu(A_i) \leqslant \frac{1}{2^i}$ and $\nu(A_i) \geqslant \varepsilon$ for every $i = 1, 2, \dots$. Let $B_j = \bigcup_{i=j}^{\infty} A_i$, $j = 1, 2, \dots$, and $B = \bigcap_{j=1}^{\infty} B_j$. Then

$$\mu(B) \leqslant \mu(B_j) = \mu\left(\bigcup_{i=j}^{\infty} A_i\right) \leqslant \sum_{i=j}^{\infty} \mu(A_j) \leqslant \sum_{i=j}^{\infty} \frac{1}{2^i} = \frac{1}{2^{j-1}} \xrightarrow{j\to\infty} 0.$$

Thus $\mu(B) = 0$ On the other hand, since $B_{j+1} \subset B_j$, $j = 1, 2, \dots$ and $\nu(\mathbb{R}^n) < \infty$, we have

$$\nu(B) = \nu\left(\bigcap_{j=1}^{\infty} B_j\right) = \lim_{j\to\infty} \nu(B_j) \geqslant \varepsilon.$$

This is a contradiction with $\nu \ll \mu$. ■

If $\nu(\mathbb{R}^n) = \infty$, then the ε, δ-version of absolute continuity implies $\nu \ll \mu$, but the converse is not tue in general. For example, let μ be the Lebesgue measure on $\mathbb{R}$ and

$$\nu(A) = \int_A \frac{1}{x}\,dx$$

for every measurable set $A \subset \mathbb{R}$. Then $\nu \ll \mu$, but the ε, δ-version of absolute continuity fails.

The following theorem on absolutely continuous measures is very important. It shows that differentiation of measures and integration are inverse operations and, in that sense, it is a version of the fundamental theorem of calculus for Radon measures. 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. Let μ and ν are Radon measures on $\mathbb{R}^n$. Recall that, by Theorem 4.19, the derivative $D_\mu \nu(x)$ exists and is finite for μ-almost every $x \in \mathbb{R}^n$.

Theorem 4.23 (Radon-Nikodym theorem). Let μ and ν are Radon measures on $\mathbb{R}^n$. Then

$$\int_A D_\mu \nu \, d\mu \leqslant \nu(A)$$

for every μ-measurable set $A \subset \mathbb{R}^n$, with equality if and only if $\nu \ll \mu$.

T E R M I N O L O G Y : We call $D_\mu \nu$ the Radon-Nikodym derivative.

T H E M O R A L : The Radon-Nikodym theorem asserts that if ν is absolutely continuous with respect to ν, then 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.

Proof. Step 1 Since μ is a Radon measure, a μ-measurable set can be written as $A = B \cup F$, where $B \subset A$ is a Borel set and $F \subset A$ with $\mu(F) = 0$. If we can show that

$$\int_B D_\mu \nu \, d\mu \leqslant \nu(B)$$

for every Borel set $B \subset \mathbb{R}^n$, then

$$\int_A D_\mu \nu \, d\mu = \int_B D_\mu \nu \, d\mu \leqslant \nu(B) \leqslant \nu(A).$$

Thus we may assume that A is a Borel set.

Let $1 < t < \infty$ and

$$A_i = \{x \in A : t^i \leqslant D_\mu \nu(x) < t^{i+1}\}, \quad i \in \mathbb{Z}.$$

The sets A_i, $i = 1, 2, \dots$, are pairwise disjoint and

$$\bigcup_{i=-\infty}^{\infty} A_i = \{x \in A : 0 < D_\mu \nu(x) < \infty\}.$$

Lemma 4.16 shows that the sets A_i, $i = 1, 2, \dots$, are Borel sets and thus μ-measurable and ν-measurable. Let

$$Z = \{x \in A : D_\mu \nu(x) = 0\},$$

$$I = \{x \in A : \overline{D}_\mu \nu(x) = \infty\}$$

and

$$N = \{x \in A : \overline{D}_\mu \nu(x) \neq \underline{D}_\mu \nu(x)\}.$$

Then

$$A = \left(\bigcup_{i=-\infty}^{\infty} A_i \right) \cup (Z \cup I \cup N).$$

By Theorem 4.19, we have $\mu(I) = 0$ and $\mu(N) = 0$.

Step 2 Since $D_\mu \nu \geqslant t^i$ in A_i, Theorem 4.17 (2) implies that

$$\nu(A_i) \geqslant t^i \mu(A_i) \quad i = 1, 2, \dots.$$

Since $D_\mu \nu \leqslant t^{i+1}$ in A_i, by Chebyshev's inequality we have

$$t^{i+1} \mu(A_i) \geqslant \int_{A_i} D_\mu \nu \, d\mu, \quad i = 1, 2, \dots.$$

Thus we have

$$\begin{aligned}
\nu(A) &\geqslant \nu\Big(\bigcup_{i=-\infty}^{\infty} A_i \Big) = \sum_{i=-\infty}^{\infty} \nu(A_i) \\
&\geqslant \sum_{i=-\infty}^{\infty} t^i \mu(A_i) = \frac{1}{t} \sum_{i=-\infty}^{\infty} t^{i+1} \mu(A_i) \\
&\geqslant \frac{1}{t} \sum_{i=-\infty}^{\infty} \int_{A_i} D_\mu \nu \, d\mu = \frac{1}{t} \int_{\bigcup_{i=1}^{\infty} A_i} D_\mu \nu \, d\mu.
\end{aligned}$$

Since

$$\int_A D_\mu \nu \, d\mu = \int_{\bigcup_{i=-\infty}^{\infty} A_i} D_\mu \nu \, d\mu + \underbrace{\int_Z D_\mu \nu \, d\mu}_{=0, D_\mu \nu = 0} + \underbrace{\int_I D_\mu \nu \, d\mu}_{=0, \mu(I)=0} + \underbrace{\int_N D_\mu \nu \, d\mu}_{=0, \mu(N)=0},$$

we have

$$\nu(A) \geqslant \frac{1}{t} \int_A D_\mu \nu \, d\mu.$$

By letting $t \to 1$, we conclude that

$$\int_A D_\mu \nu \, d\mu \leqslant \nu(A)$$

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

Step 2 Assume that $\nu \ll \mu$ and that $A \subset \mathbb{R}^n$ is a μ-measurable set. Since μ is a Radon measure, a μ-measurable set can be written as $A = B \cup F$, where $B \subset A$ is a Borel set and $F \subset A$ with $\mu(F) = 0$. Since $\nu \ll \mu$ we have $\nu(F) = 0$. Thus $A = B \cup F$, where $B \subset A$ is a Borel set and $F \subset A$ with $\nu(F) = 0$. This implies that A is ν-measurable.

Step 3 By Theorem 4.19, we have $\mu(I) = 0$ and $\mu(N) = 0$. Since $\nu \ll \mu$, we have $\nu(I) = 0$ and $\nu(N) = 0$.

Let $0 < t < \infty$. Since $D_\mu \nu = 0 \leqslant t$ in Z, Theorem 4.17 (1) implies that

$$\nu(Z \cap B(0,i)) \leqslant t \underbrace{\mu(Z \cap B(0,i))}_{<\infty}, \quad i = 1, 2, \dots.$$

By letting $t \to 0$, we conclude that $\nu(Z \cap B(0,i)) = 0$ for every $i = 1, 2, \dots$. Thus

$$\nu(Z) = \nu\Big(\bigcup_{i=1}^{\infty} (Z \cap B(0,i))\Big) \leqslant \sum_{i=1}^{\infty} \underbrace{\nu(Z \cap B(0,i))}_{=0} = 0.$$

This shows that $\nu(Z) = 0$.

Since $A \setminus \bigcup_{i=-\infty}^{\infty} A_i = Z \cup I \cup N$, we have

$$\nu\Big(A \setminus \bigcup_{i=-\infty}^{\infty} A_i\Big) \leqslant \nu(Z \cup I \cup N) \leqslant \nu(Z) + \nu(I) + \nu(N) = 0.$$

Step 4 Since $D_\mu \nu \leqslant t^{i+1}$ in A_i, Theorem 4.17 (1) implies that

$$\nu(A_i) \leqslant t^{i+1} \mu(A_i), \quad i = 1, 2, \dots.$$

Since $D_\mu \nu \geqslant t^i$ in A_i, by Chebyshev's inequality we have

$$t^i \mu(A_i) \leqslant \int_{A_i} D_\mu \nu \, d\mu, \quad i = 1, 2, \dots.$$

Thus we have

$$\begin{aligned}\nu(A) &= \nu\Big(A\cap\bigcup_{i=-\infty}^{\infty}A_i\Big)+\underbrace{\nu\Big(A\setminus\bigcup_{i=-\infty}^{\infty}A_i\Big)}_{=0}=\sum_{i=-\infty}^{\infty}\nu(A_i)\\ &\leqslant \sum_{i=-\infty}^{\infty}t^{i+1}\mu(A_i)=t\sum_{i=-\infty}^{\infty}t^i\mu(A_i)\\ &\leqslant t\sum_{i=-\infty}^{\infty}\int_{A_i}D_\mu\nu\,d\mu=t\int_{\bigcup_{i=-\infty}^{\infty}A_i}D_\mu\nu\,d\mu\leqslant t\int_A D_\mu\nu\,d\mu.\end{aligned}$$

By letting $t\to 1$ we arrive at

$$\nu(A)\leqslant\int_A D_\mu\nu\,d\mu.$$

Step 4 If

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

for every μ-measurable set $A\subset\mathbb{R}^n$, then $\nu\ll\mu$. This follows, since $\nu(A)=0$ if $\mu(A)=0$. □

Remarks 4.24:

(1) It can be shown that $\nu\ll\mu$ if and only of $D_\mu\nu(x)<\infty$ for μ almost every $x\in\mathbb{R}^n$ (exercise). This is a pointwise characterization of absolute continuity of measures.

(2) By Example 4.21 (2) we may conclude that $\nu\ll\mu$ if and only if $\nu(A)=\int_A f\,d\mu$, where f is a nonnegative μ-measurable function A is a μ-measurable set.

(3) The Radon-Nikodym derivative is unique: If $f\in L^1_{\mathrm{loc}}(\mathbb{R}^n;\mu)$ is a nonnegative function and $\nu(A)=\int_A f\,d\mu$ for every μ-measurable set $A\subset\mathbb{R}^n$, then $f=D_\mu\nu$ μ-almost everywhere (exercise).

(4) 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 $\nu(\mathbb{R}^n)<\infty$ (exercise).

Remark 4.25. 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

$$\int_A f\,d\nu=\int_A fD_\mu\nu\,d\mu$$

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

$$\int_A f\,d\nu=\int_A fg\,d\mu.$$

(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 4.26. 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 4.27. 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.

Example 4.28. 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 $\mu(\{0\}) = \nu(\mathbb{R}^n \setminus \{0\}) = 0$. Thus $\nu \perp \mu$.

Remark 4.29. Absolutely continuous and singular measures have the following 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 4.30 (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

$$\nu(A) = \int_A D_\mu \nu \, d\mu + \nu_s(A)$$

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

T E R M I N O L O G Y : We call ν_a the absolutely continuous part and ν_s the singular part of ν with respect to μ.

T H E M O R A L : Any Radon measure can be split into absolutely continuous and singular parts with respect to another Radon measure. The absolutely continuous part can be represented as an integral of the derivative of the measures. Moreover, the absolutely continuous part lives in the set where $\underline{D}_\mu \nu < \infty$ and the singular part in the set where $\underline{D}_\mu \nu = \infty$.

Proof. Step 1 Let

$$B = \{x \in \mathbb{R}^n : \underline{D}_\mu \nu(x) < \infty\},$$

$$\nu_a = \nu \lfloor B \quad \text{and} \quad \nu_s = \nu \lfloor (\mathbb{R}^n \setminus B).$$

Here $\lfloor$ denotes the restriction of a measure to a subset. Then $\nu = \nu_a + \nu_s$ and, by the properties of restrictions of measures, ν_a and ν_s are Radon measures.

Step 2 By Theorem 4.19, the derivative $D_\mu \nu(x)$ exists and is finite for μ-almost every $x \in \mathbb{R}^n$. Thus

$$\mu(\mathbb{R}^n \setminus B) = 0 = \nu(\emptyset) = \mu(B \cap (\mathbb{R}^n \setminus B)) = \nu_s(B).$$

This shows that $\nu_s \perp \mu$.

Step 3 Let $A \subset \mathbb{R}^n$ with $\mu(A) = 0$. Let

$$B_i = \left\{x \in \mathbb{R}^n : D_\mu \nu(x) \leqslant i\right\}, \quad i = 1, 2, \dots.$$

Then $B_i \subset B_{i+1}$, $i = 1, 2, \dots$, and $B = \bigcup_{i=1}^\infty B_i$. Since $D_\mu \nu \leqslant i$ in $A \cap B_i$, Theorem 4.17 (1) implies that

$$\nu(A \cap B_i) \leqslant i\mu(A \cap B_i) \leqslant i\mu(A) = 0, \quad i = 1, 2, \dots.$$

It follows that

$$\begin{aligned} \nu_a(A) = \nu(A \cap B) &= \nu\left(A \cap \bigcup_{i=1}^\infty B_i\right) = \nu\left(\bigcup_{i=1}^\infty (A \cap B_i)\right) \\ &\leqslant \sum_{i=1}^\infty \nu(A \cap B_i) = 0. \end{aligned}$$

This shows that $\nu_a \ll \mu$. Theorem 4.23 implies that

$$\nu_a(A) = \int_A D_\mu \nu_a \, d\mu$$

Step 4 Let

$$C_i = \left\{x \in \mathbb{R}^n : D_\mu \nu_s(x) \geqslant \tfrac{1}{i}\right\}, \quad i = 1, 2, \dots.$$

Since $D_\mu \nu_s \geqslant \frac{1}{i}$ in $C_i \cap B$, Theorem 4.17 (2) implies that

$$\nu_s(C_i \cap B) \geqslant \tfrac{1}{i}\mu(C_i \cap B), \quad i = 1, 2, \dots.$$

Then

$$\begin{aligned}\tfrac{1}{i}\mu(C_i) &= \tfrac{1}{i}(\mu(C_i \cap B) + \underbrace{\mu(C_i \cap (\mathbb{R}^n \setminus B))}_{=0,\,\mu(\mathbb{R}^n \setminus B)=0} \\ &\leqslant \nu_s(C_i \cap B) = \nu(\underbrace{(C_i \cap B) \cap (\mathbb{R}^n \setminus B)}_{=\emptyset}) = 0, \quad i = 1, 2, \dots.\end{aligned}$$

This shows that $\mu(C_i) = 0$ for every $i = 1, 2, \dots$, 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

$$\begin{aligned}D_\mu \nu_a(x) &= \lim_{r\to 0} \frac{\nu_a(B(x,r))}{\mu(B(x,r))} = \lim_{r\to 0} \frac{\nu((B(x,r)) - \nu_s(B(x,r))}{\mu(B(x,r))} \\ &= D_\mu \nu(x) - D_\mu \nu_s(x) = D_\mu \nu(x)\end{aligned}$$

for μ-almost every $x \in \mathbb{R}^n$. Since $\nu_a \ll \mu$, the Radon-Nikodym theorem (Theorem 4.23) implies that

$$\nu_a(A) = \int_A D_\mu \nu_a \, d\mu$$

for every μ-measurable set $A \subset \mathbb{R}^n$.

$$\begin{aligned}\nu(A) &= \nu_a(A) + \nu_s(A) \\ &= \int_A D_\mu \nu_a \, d\mu + \nu_s(A) \\ &= \int_A D_\mu \nu \, d\mu + \nu_s(A).\end{aligned}$$

□

Remarks 4.31:

(1) It can be shown that $\mu \perp \nu$ if and only if $D_\mu \nu(x) = 0$ for μ almost every $x \in \mathbb{R}^n$ (exercise). This is a pointwise characterization of mutual singularity of measures.

(2) $\nu \ll \mu$ if and only if $\nu_s = 0$ (exercise).

Remark 4.32. 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 $\mathbb{R}^n$, see Theorem 2.24 for the case of the Lebesgue measure.

Theorem 4.33 (Lebesgue differentiation theorem). Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n;\mu)$. 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. Let

$$\nu^{\pm}(B) = \int_B f^{\pm}\,d\mu,$$

where $B \subset \mathbb{R}^n$ is a Borel set, and

$$\nu^{\pm}(A) = \inf\{\nu^{\pm}(B) : A \subset B,\ B \text{ Borel}\}$$

for an arbitrary set $A \subset \mathbb{R}^n$. Then ν^+ and ν^- are Radon measures and $\nu^{\pm} \ll \mu$ (exercise). The Radon-Nikodym theorem (Theorem 4.23) 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,

$$\begin{aligned}
\lim_{r\to 0}\frac{1}{\mu(B(x,r))}\int_{B(x,r)} f\,d\mu &= \lim_{r\to 0}\frac{1}{\mu(B(x,r))}\left(\int_{B(x,r)} f^+\,d\mu - \int_{B(x,r)} f^-\,d\mu\right)\\
&= \lim_{r\to 0}\frac{1}{\mu(B(x,r))}\big(\nu^+(B(x,r)) - \nu^-(B(x,r))\big)\\
&= D_\mu\nu^+(x) - D_\mu\nu^-(x)\\
&= f^+(x) - f^-(x) = f(x)
\end{aligned}$$

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

Corollary 4.34. Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n;\mu)$. 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 4.33, for every $i = 1,2,\dots$ 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

$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - q_i| \, d\mu = |f(x) - q_i| \quad \text{for every} \quad x \in \mathbb{R}^n \setminus A.$$

Let $x \in \mathbb{R}^n \setminus A$ and $\varepsilon > 0$. Then there exists q_i such that $|f(x) - q_i| < \frac{\varepsilon}{2}$. This implies

$$\begin{aligned}
&\limsup_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - f(x)| \, d\mu \\
&\quad \leqslant \limsup_{r\to 0} \frac{1}{\mu(B(x,r))} \left(\int_{B(x,r)} |f - q_i| \, d\mu + \int_{B(x,r)} |q_i - f(x)| \, d\mu \right) \\
&\quad = |f(x) - q_i| + |f(x) - q_i| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

□

Remarks 4.35:

(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 $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n; \mu)$ 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 $\mathbb{R}^n$.

It is also possible to consider the centered maximal function $M_\mu f : \mathbb{R}^n \to [0,\infty]$ of $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n;\mu)$ associated with μ defined by

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

The Hardy-Littlewood-Wiener maximal function theorems hold for the maximal operator with respect to a general Radon measure, compare with Theorem 2.17 and Theorem 2.22. This gives an alternative approach to prove the Lebesgue differentiation theorem, Theorem 4.33, for a general Radon measure as in the proof of Theorem 2.24.

Theorem 4.36 (General maximal function theorem). Let μ be a Radon measure on $\mathbb{R}^n$. There exists a constant $c = c(n)$ such that

$$\mu(\{x \in \mathbb{R}^n : M_\mu f(x) > \lambda\}) \leqslant \frac{c}{\lambda} \|f\|_{L^1(\mathbb{R}^n;\mu)} \quad \text{for every} \quad \lambda > 0.$$

For $1 < p < \infty$, there exists a constant $c = c(n,p)$ such that

$$\|Mf\|_{L^p(\mathbb{R}^n;\mu)} \leqslant c\|f\|_{L^p(\mathbb{R}^n;\mu)}.$$

Proof. The proof of the weak type estimate is similar to the proof of Theorem 2.17, but instead of the covering lemma, see Theorem 2.15, we use Besicovitch covering theorem, Theorem 4.2. For every point $x \in \mathbb{R}^n$ with $M_\mu f(x) > \lambda$ there exists a ball centered at x for which the average appearing in the definition of the maximal function is greater than λ. The constant appearing in the weak type estimate is the constant P in Theorem 4.2. The strong type estimate follows by applying the weak type estimate and the $L^\infty(\mathbb{R}^n;\mu)$ estimate as in the proof of 2.22 (exercise). □

Remark 4.37. The noncentered version of the maximal function M_μ satisfies similar weak type and strong type inequalities in the one-dimensional case. In the higher dimensional case these results do not hold in general. However, if the measure μ is doubling, the noncentered operator satisfies weak type and strong type estimates. In this case we can apply the covering lemma, see Theorem 2.15, and the constants in depend on the doubling constant of the measure.

THE MORAL: The Hardy-Littlewood-Wiener maximal function theorems hold for the centered maximal operator with respect to a general Radon measure, but we have to use a more powerfull covering theorem compared to the Lebesgue measure.

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 Weak convergence methods

Radon measures on $\mathbb{R}^n$ interact nicely with the Euclidean topology. Indeed, measurable sets can be approximated by 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

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 on $\mathbb{R}^n$. 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, $\|L\| = \|g\|_{p'}$.

THE MORAL: This shows that for every function $g \in L^{p'}(\mathbb{R}^n)$ there exists 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 $\|g\|_{p'} = 0$, then $L(f) = 0$ for every $f \in L^p(\mathbb{R}^n)$ and the claim is clear. Hence we may assume that $\|g\|_{p'} > 0$.

$\boxed{1 < p < \infty}$ By Hölder's inequality

$$\begin{aligned} |L(f)| &= \left| \int_{\mathbb{R}^n} fg \, d\mu \right| \leqslant \int_{\mathbb{R}^n} |f||g| \, d\mu \\ &\leqslant \left(\int_{\mathbb{R}^n} |f|^p \, d\mu \right)^{\frac{1}{p}} \left(\int_{\mathbb{R}^n} |g|^{p'} \, d\mu \right)^{\frac{1}{p'}} = \|f\|_p \|g\|_{p'}. \end{aligned}$$

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 $f = |g|^{\frac{p'}{p}} \operatorname{sign} g$ belongs to $L^p(\mathbb{R}^n)$, since

$$\|f\|_p = \left(\int_{\mathbb{R}^n} |f|^p \, d\mu \right)^{\frac{1}{p}} = \left(\int_{\mathbb{R}^n} |g|^{p'} \, d\mu \right)^{\frac{1}{p}} = \|g\|_{p'}^{\frac{p'}{p}} < \infty.$$

Since $|g|^{\frac{p'}{p}} g \operatorname{sign} g = |g|^{\frac{p'}{p}} |g| = |g|^{p'} \geqslant 0$, we have

$$\begin{aligned} |L(f)| = L(f) &= \int_{\mathbb{R}^n} |g|^{\frac{p'}{p}} \underbrace{g \operatorname{sign} g}_{=|g|} \, d\mu = \int_{\mathbb{R}^n} |g|^{p'} \, d\mu \\ &= \left(\int_{\mathbb{R}^n} |g|^{p'} \, d\mu \right)^{\frac{1}{p'}} \left(\int_{\mathbb{R}^n} \underbrace{|g|^{p'}}_{=|f|^p} \, d\mu \right)^{\frac{1}{p}} = \|f\|_p \|g\|_{p'} \end{aligned}$$

and
$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \neq 0} \frac{|L(f)|}{\|f\|_p} \geqslant \|g\|_{p'}.$$
This shows that $\|L\| = \|g\|_{p'}$.

$\boxed{p=\infty}$ Assume that $g \in L^1(\mathbb{R}^n)$. Again we may assume that $\|g\|_1 > 0$. Then
$$|L(f)| \leqslant \int_{\mathbb{R}^n} |f||g|\,d\mu \leqslant \|f\|_\infty \|g\|_1,$$
which implies that
$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_\infty \leqslant 1} |L(f)| \leqslant \|g\|_1 < \infty.$$
On the other hand, since the function $f = \operatorname{sign} g$ belongs to $L^\infty(\mathbb{R}^n)$, $\|f\|_\infty \leqslant 1$ and $g \operatorname{sign} g = |g| \geqslant 0$, we have
$$|L(f)| = L(f) = \int_{\mathbb{R}^n} g \operatorname{sign} g\,d\mu = \int_{\mathbb{R}^n} |g|\,d\mu = \|g\|_1.$$
This shows that
$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_\infty \leqslant 1} |L(f)| \geqslant \|g\|_1,$$
from which it follows that $\|L\| = \|g\|_1$.

$\boxed{p=1}$ Let $g \in L^\infty(\mathbb{R}^n)$. Again we may assume that $\|g\|_\infty > 0$. Then
$$|L(f)| \leqslant \int_{\mathbb{R}^n} |f||g|\,d\mu \leqslant \|f\|_1 \|g\|_\infty,$$
which implies that
$$\|L\| = \sup_{f \in L^1(\mathbb{R}^n), \|f\|_1 \leqslant 1} |L(f)| \leqslant \|g\|_\infty < \infty.$$
Assume first that $\mu(\mathbb{R}^n) < \infty$. Let $0 < \varepsilon < \|g\|_\infty$,
$$A_\varepsilon = \{x \in \mathbb{R}^n : |g(x)| \geqslant \|g\|_\infty - \varepsilon\} \quad \text{and} \quad f_\varepsilon = \frac{\chi_{A_\varepsilon} \operatorname{sign} g}{\mu(A_\varepsilon)}.$$
Then $0 < \mu(A_\varepsilon) \leqslant \mu(\mathbb{R}^n) < \infty$. We observe that $f_\varepsilon \in L^1(\mathbb{R}^n)$ and
$$\|f_\varepsilon\|_1 = \int_{\mathbb{R}^n} \left| \frac{\chi_{A_\varepsilon} \operatorname{sign} g}{\mu(A_\varepsilon)} \right| d\mu \leqslant \int_{\mathbb{R}^n} \frac{\chi_{A_\varepsilon}}{\mu(A_\varepsilon)}\,d\mu = \frac{\mu(A_\varepsilon)}{\mu(A_\varepsilon)} = 1.$$
Thus
$$|L(f_\varepsilon)| = \left| \int_{\mathbb{R}^n} f_\varepsilon g\,d\mu \right| = \int_{A_\varepsilon} \frac{|g|}{\mu(A_\varepsilon)}\,d\mu \geqslant \int_{A_\varepsilon} \frac{\|g\|_\infty - \varepsilon}{\mu(A_\varepsilon)}\,d\mu = \|g\|_\infty - \varepsilon.$$
This shows that, for every $0 < \varepsilon < \|g\|_\infty$, there exists $f_\varepsilon \in L^1(\mathbb{R}^n)$ with $\|f_\varepsilon\|_1 \leqslant 1$ such that
$$\|g\|_\infty - \varepsilon \leqslant |L(f_\varepsilon)| \leqslant \|g\|_\infty.$$

This implies

$$\|L\| = \sup_{f\in L^1(\mathbb{R}^n),\|f\|_1\leqslant 1} |L(f)| = \|g\|_\infty$$

under the assumption that $\mu(\mathbb{R}^n) < \infty$.

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 $i = 1,2,\dots$. For example, we may choose $A_i = B(0,i)$, $i = 1,2,\dots$. Let

$$A_{\varepsilon,i} = \{x \in B(0,i) : |g(x)| \geqslant \|g\|_\infty - \varepsilon\} \quad \text{and} \quad f_{\varepsilon,i} = \frac{\chi_{A_{\varepsilon,i}} \operatorname{sign} g}{\mu(A_{\varepsilon,i})}$$

for $i = 1,2,\dots$. We observe that $A_{\varepsilon,i} \subset A_{\varepsilon,i+1}$, $i = 1,2,\dots$, and $A_\varepsilon = \bigcup_{i=1}^\infty A_{\varepsilon,i}$. Since $\mu(A_\varepsilon) > 0$ and we have

$$0 < \mu(A_\varepsilon) = \mu\left(\bigcup_{i=1}^\infty A_{\varepsilon,i}\right) = \lim_{i\to\infty} \mu(A_{\varepsilon,i})$$

and, consequently, there exists i such that $\mu(A_{\varepsilon,i}) > 0$. On the other hand, $\mu(A_{\varepsilon,i}) \leqslant \mu(B(0,i)) < \infty$. As above, we conclude that, for every $\varepsilon > 0$, there exists $f_{\varepsilon,i} \in L^1(\mathbb{R}^n)$ with $\|f_{\varepsilon,i}\|_1 \leqslant 1$ such that

$$\|g\|_\infty - \varepsilon \leqslant |L(f_{\varepsilon,i})| \leqslant \|g\|_\infty.$$

This implies

$$\|L\| = \sup_{f\in L^1(\mathbb{R}^n),\|f\|_1\leqslant 1} |L(f)| = \|g\|_\infty. \qquad \square$$

Then we show that the converse of the previous theorem holds for $1 \leqslant p < \infty$.

Theorem 5.2 (Riesz representation theorem in L^p). Let $1 \leqslant p < \infty$ and assume that μ is a Radon measure. For every bounded linear functional $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ there exists a unique $g \in L^{p'}(\mathbb{R}^n)$ such that

$$L(f) = \int_{\mathbb{R}^n} f g \, d\mu \quad \text{for every} \quad f \in L^p(\mathbb{R}^n). \tag{5.3}$$

Moreover, $\|L\| = \|g\|_{p'}$.

THE MORAL: The dual space of $L^p(\mathbb{R}^n)$ is isomorphic to $L^{p'}(\mathbb{R}^n)$, that is, $L^p(\mathbb{R}^n)^* = L^{p'}(\mathbb{R}^n)$ for $1 \leqslant p < \infty$.

WARNING: 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. If $\|L\| = 0$, then $g = 0$ in $L^{p'}(\mathbb{R}^n)$, that is $g = 0$ μ-almost everywhere in $\mathbb{R}^n$, satisfies the required properties. Thus we may assume that $\|L\| > 0$. First we assume that L is a positive functional, that is, $f \geqslant 0$ μ-almost everywhere in $\mathbb{R}^n$ implies $L(f) \geqslant 0$ At the end of the proof, we show that every bounded linear functional $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ can be represented as a difference of two positive functionals.

(1) First we assume that $\mu(\mathbb{R}^n) < \infty$. Later in the proof we show that this assumption can be removed for σ-finite measures. For a μ-measurable set $A \subset \mathbb{R}^n$, let

$$\nu(A) = L(\chi_A).$$

We have

$$\int_{R^n} |\chi_A|^p \, d\mu = \mu(A) \leqslant \mu(\mathbb{R}^n) < \infty,$$

which implies that $\chi_A \in L^p(\mathbb{R}^n)$ and ν is well defined. Observe that $\chi_A \geqslant 0$ and since the operator is positive, we have $\nu(A) = L(\chi_A) \geqslant 0$. Thus ν is a nonnegative set function on μ-measurable sets.

(2) CLAIM: ν is countably additive on pairwise disjoint μ-measurable sets.

Reason. Assume that A_i, $i = 1, 2, \dots$, are pairwise disjoint μ-measurable sets. Let $B = \bigcup_{i=1}^{\infty} A_i$ and $B_k = \bigcup_{i=1}^{k} A_i$. Since $B_k \subset B_{k+1}$ and B_k is a μ-measurable set, $k = 1, 2, \dots$, we have

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

This implies

$$\|\chi_B - \chi_{B_k}\|_p^p = \int_{\mathbb{R}^n} |\chi_B - \chi_{B_k}|^p \, d\mu = \mu(B \setminus B_k) = \mu(B) - \mu(B_k) \xrightarrow{k\to\infty} 0,$$

since $\mu(B) \leqslant \mu(\mathbb{R}^n) < \infty$. It follows that $\chi_{B_k} \to \chi_B$ in $L^p(\mathbb{R}^n)$ and by the continuity of L, we have $L(\chi_{B_k}) \to L(\chi_B)$ as $k \to \infty$. This implies that

$$\begin{aligned}\sum_{i=1}^{\infty} \nu(A_i) &= \lim_{k\to\infty} \sum_{i=1}^{k} \nu(A_i) = \lim_{k\to\infty} \nu(B_k) \\ &= \lim_{k\to\infty} L(\chi_{B_k}) = L(\chi_B) = \nu(B) = \nu\left(\bigcup_{i=1}^{\infty} A_i\right).\end{aligned}$$

■

(3) CLAIM: ν is absolutely continuous with respect to μ.

Reason. If $\mu(A) = 0$, then $\|\chi_A\|_p = 0$ and thus $\chi_A = 0$ in $L^p(\mathbb{R}^n)$. Since a linear functional maps zero to zero, we have $\nu(A) = L(\chi_A) = 0$. ■

(4) By the Radon-Nikodym theorem, see Theorem 4.23, 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 (5.3) for the characteristic functions $f = \chi_A$, where A is a μ-measurable set. We still have to show that $g \in L^{p'}(\mathbb{R}^n)$, (5.3) holds for all $f \in L^p(\mathbb{R}^n)$ and $\|L\| = \|g\|_{p'}$.

(5) CLAIM: The representation in (5.3) holds for every $f \in L^{\infty}(\mathbb{R}^n)$ with $f \geqslant 0$ μ-almost everywhere in $\mathbb{R}^n$.

Reason. In (4) we showed that (5.3) holds for every characteristic function of a μ-measurable set and consequently it holds for linear combinations of such sets. Thus (5.3) holds for simple functions. Every nonnegative $f \in L^\infty(\mathbb{R}^n)$ can be approximated in $L^\infty(\mathbb{R}^n)$ by an increasing sequence (s_i) of nonnegative simple functions. Since $\mu(\mathbb{R}^n) < \infty$, we have

$$\|s_i - f\|_p = \left(\int_{\mathbb{R}^n} |s_i - f|^p \, d\mu\right)^{\frac{1}{p}} \leqslant \|s_i - f\|_\infty \mu(\mathbb{R}^n)^{\frac{1}{p}} \xrightarrow{i\to\infty} 0$$

and since L is a bounded operator, we have

$$|L(s_i) - L(f)| \leqslant \|L\| \|s_i - f\|_p \xrightarrow{i\to\infty} 0.$$

On the other hand,

$$\begin{aligned}\left|\int_{\mathbb{R}^n} s_i g \, d\mu - \int_{\mathbb{R}^n} f g \, d\mu\right| &\leqslant \int_{\mathbb{R}^n} |s_i - f||g| \, d\mu \\ &\leqslant \|s_i - f\|_\infty \int_{\mathbb{R}^n} |g| \, d\mu \xrightarrow{i\to\infty} 0.\end{aligned}$$

Thus

$$L(f) = \lim_{i\to\infty} L(s_i) = \lim_{i\to\infty} \int_{\mathbb{R}^n} s_i g \, d\mu = \int_{\mathbb{R}^n} f g \, d\mu.$$

This shows (5.3) holds for every $f \in L^\infty(\mathbb{R}^n)$ with $f \geqslant 0$ μ-almost everywhere in $\mathbb{R}^n$. ■

(6) For a general sign-changing function $f \in L^\infty(\mathbb{R}^n)$, we represent f as a difference of the positive and negative parts $f = f^+ - f^-$, where $f^+ \in L^\infty(\mathbb{R}^n)$ and $f^- \in L^\infty(\mathbb{R}^n)$. We apply (5) to conclude that there exists nonnegative $g \in L^1(\mathbb{R}^n)$ such that

$$\begin{aligned}L(f) &= L(f^+ - f^-) = L(f^+) - L(f^-) \\ &= \int_{\mathbb{R}^n} f^+ g \, d\mu - \int_{\mathbb{R}^n} f^- g \, d\mu \\ &= \int_{\mathbb{R}^n} (f^+ - f^-) g \, d\mu \\ &= \int_{\mathbb{R}^n} f g \, d\mu.\end{aligned}$$

This shows (5.3) holds for every $f \in L^\infty(\mathbb{R}^n)$

(7) CLAIM: $\|g\|_{p'} \leqslant \|L\|$ and thus $g \in L^{p'}(\mathbb{R}^n)$. Recall that $g \geqslant 0$ with $\|g\|_{p'} \neq 0$.

Reason. $1 < p < \infty$ Let

$$A_i = \{x \in \mathbb{R}^n : g(x) \leqslant i\} \quad \text{and} \quad f_i = \chi_{A_i} g^{p'-1}, \quad i = 1, 2, \ldots.$$

Then $f_i \in L^\infty(\mathbb{R}^n)$ and $f_i^p = g^{p'}$ on A_i. Thus

$$\int_{A_i} g^{p'}\,d\mu = \int_{\mathbb{R}^n} f_i g\,d\mu = L(f_i) \leqslant \|L\|\|f_i\|_p = \|L\|\left(\int_{A_i} g^{p'}\,d\mu\right)^{\frac{1}{p}}.$$

This implies

$$\left(\int_{A_i} g^{p'}\,d\mu\right)^{\frac{1}{p'}} = \left(\int_{A_i} g^{p'}\,d\mu\right)^{1-\frac{1}{p}} \leqslant \|L\|$$

for every $i = 1,2,\dots$. By the monotone convergence theorem, we have

$$\begin{aligned}
\int_{\mathbb{R}^n} g^{p'}\,d\mu &= \int_{\bigcup_{i=1}^\infty A_i} g^{p'}\,d\mu = \int_{\mathbb{R}^n} g^{p'}\chi_{\bigcup_{i=1}^\infty A_i}\,d\mu \\
&= \int_{\mathbb{R}^n} g^{p'} \lim_{i\to\infty} \chi_{A_i}\,d\mu = \lim_{i\to\infty}\int_{\mathbb{R}^n} g^{p'}\chi_{A_i}\,d\mu \\
&= \lim_{i\to\infty}\int_{A_i} g^{p'}\,d\mu \leqslant \|L\|^{p'}.
\end{aligned}$$

This shows that $\|g\|_{p'} \leqslant \|L\|$.

$\boxed{p = 1,\, p' = \infty}$ Let

$$A_i = \left\{x \in \mathbb{R}^n : g(x) \geqslant \|L\| + \tfrac{1}{i}\right\}, \quad i = 1,2,\dots.$$

Then

$$\begin{aligned}
\left(\|L\| + \tfrac{1}{i}\right)\mu(A_i) &\leqslant \int_{A_i} g\,d\mu = \left|\int_{\mathbb{R}^n} \chi_{A_i} g\,d\mu\right| \\
&= |L(\chi_{A_i})| = \|L\|\|\chi_{A_i}\|_1 \leqslant \|L\|\mu(A_i),
\end{aligned}$$

which can happen only if $\mu(A_i) = 0$ for every $i = 1,2,\dots$. 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 $\|g\|_\infty \leqslant \|L\|$. ■

$\boxed{(8)}$ Thus $g \in L^{p'}(\mathbb{R}^n)$ and

$$L(f) = \int_{\mathbb{R}^n} fg\,d\mu$$

for every $f \in L^\infty(\mathbb{R}^n)$. Both sides of the equality above 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 the equality above holds for every $f \in L^p(\mathbb{R}^n)$ under the assumption $\mu(\mathbb{R}^n) < \infty$.

To show the uniqueness of g, assume that there exist $g_1, g_2 \in L^{p'}(\mathbb{R}^n)$ such that

$$L(f) = \int_{\mathbb{R}^n} fg_1\,d\mu \quad \text{and} \quad L(f) = \int_{\mathbb{R}^n} fg_2\,d\mu$$

for every $f \in L^p(\mathbb{R}^n)$. It follows that

$$\int_{\mathbb{R}^n} f(g_1 - g_2)\, d\mu = 0$$

for every $f \in L^p(\mathbb{R}^n)$. We choose $f = \operatorname{sign}(g_1 - g_2)$. Since $\mu(\mathbb{R}^n) < \infty$, we have

$$\int_{\mathbb{R}^n} |f|^p \, d\mu = \int_{\mathbb{R}^n} |\operatorname{sign}(g_1 - g_2)|^p \, d\mu \leqslant \mu(\mathbb{R}^n) < \infty.$$

Thus $f \in L^p(\mathbb{R}^n)$. It follows that

$$\int_{\mathbb{R}^n} |g_1 - g_2|\, d\mu = \int_{\mathbb{R}^n} \operatorname{sign}(g_1 - g_2)(g_1 - g_2)\, d\mu = \int_{\mathbb{R}^n} f(g_1 - g_2)\, d\mu = 0.$$

This implies that $g_1(x) = g_2(x)$ for μ-almost every $x \in \mathbb{R}^n$.

(9) CLAIM: $\|g\|_{p'} = \|L\|$.

Reason. By (7) we have $\|g\|_{p'} \leqslant \|L\|$. The opposite direction comes from Hölder's inequality, since

$$|L(f)| \leqslant \left| \int_{\mathbb{R}^n} f g \, d\mu \right| \leqslant \|f\|_p \|g\|_{p'}$$

so that

$$\|L\| = \sup_{f \in L^p(\mathbb{R}^n), \|f\|_p \leqslant 1} |L(f)| \leqslant \|g\|_{p'}.$$

■

(10) The proof is now complete in the case $\mu(\mathbb{R}^n) < \infty$. Next we consider the case $\mu(\mathbb{R}^n) = \infty$. Let

$$\mathscr{F} = \{A \subset \mathbb{R}^n : A \ \mu\text{-measurable and } \mu(A) < \infty\}.$$

Note that $\mathscr{F}$ is not a σ-algebra, since $\mathbb{R}^n \setminus A$ does not necessarily belong to $\mathscr{F}$, if $A \in \mathscr{F}$. For $A \in \mathscr{F}$, we may identify $L^p(A)$ with $\{f \in L^p(\mathbb{R}^n) : f = 0 \text{ in } \mathbb{R}^n \setminus A\}$ by extending all functions in $L^p(A)$ by zero to $\mathbb{R}^n \setminus A$. Since $\mu(A) < \infty$, we may apply the beginning of the proof to the bounded nonnegative linear functional $L : L^p(A) \to \mathbb{R}$ and obtain a unique $g_A \in L^{p'}(A)$ such that

$$L(f) = \int_A f g_A \, d\mu = \int_{\mathbb{R}^n} f g_A \, d\mu \quad \text{for every} \quad f \in L^p(A)$$

and

$$\begin{aligned} \|g_A\|_{L^{p'}(A)} &= \sup_{f \in L^p(A), \|f\|_{L^p(A)} \neq 0} \frac{|L(f)|}{\|f\|_{L^p(A)}} \\ &= \sup_{f \in L^p(A), \|f\|_{L^p(A)} \neq 0} \frac{|L(f)|}{\|f\|_{L^p(\mathbb{R}^n)}} \leqslant \|L\|. \end{aligned}$$

Extend g_A by zero to $\mathbb{R}^n \setminus A$. Since the right-hand side does not depend on $A \in \mathscr{F}$, we obtain

$$\sup_{A \in \mathscr{F}} \|g_A\|_{L^{p'}(\mathbb{R}^n)} \leqslant \|L\| < \infty.$$

By the definition of supremum, there exists a sequence of sets (A_i), $A_i \in \mathscr{F}$, $i = 1,2,\ldots$, such that

$$\lim_{i\to\infty} \|g_{A_i}\|_{L^{p'}(\mathbb{R}^n)} = \sup_{A\in\mathscr{F}} \|g_A\|_{L^{p'}(\mathbb{R}^n)}.$$

(11) CLAIM: We may assume that $A_i \subset A_{i+1}$ and $0 \leqslant g_{A_i} \leqslant g_{A_{i+1}}$ μ-almost everywhere for $i = 1,2,\ldots$.

Reason. Assume that $A,B \in \mathscr{F}$. By the uniqueness of g_A, we have $g_A(x) = g_{A\cap B}(x) = g_B(x)$ for μ-almost every $x \in A\cap B$. In particular, this implies that

$$g_{A\cup B} = \max\{g_A, g_B\}.$$

On the other, if $A \subset B$, then $0 \leqslant g_A \leqslant g_B$ μ-almost everywhere. The claim follows by replacing A_i with $\bigcup_{j=1}^{i} A_j$ and g_{A_i} with $g_{\bigcup_{j=1}^{i} A_j}$. ∎

Since (g_{A_i}) is an increasing sequence, it converges μ-almost everywhere and we may define

$$g(x) = \lim_{i\to\infty} g_{A_i}(x)$$

for μ-almost every $x \in \mathbb{R}^n$. By the monotone convergence theorem, we have

$$\begin{aligned}\|g\|_{L^{p'}(\mathbb{R}^n)} &= \lim_{i\to\infty} \|g_{A_i}\|_{L^{p'}(\mathbb{R}^n)}\\ &= \sup_{A\in\mathscr{F}} \|g_A\|_{L^{p'}(\mathbb{R}^n)} \leqslant \|L\| < \infty.\end{aligned}$$

(12) CLAIM: If $A \in \mathscr{F}$, then $g_A = g$ μ-almost everywhere in $\mathbb{R}^n$

Reason. By (10), we have

$$\begin{aligned}\sup_{A\in\mathscr{F}} \|g_A\|_{L^{p'}(\mathbb{R}^n)} = \|g\|_{L^{p'}(\mathbb{R}^n)} &= \lim_{i\to\infty} \|g_{A_i}\|_{L^{p'}(\mathbb{R}^n)}\\ &\leqslant \lim_{i\to\infty} \|g_{A_i\cup A}\|_{L^{p'}(\mathbb{R}^n)}\\ &= \lim_{i\to\infty} \|\max\{g_{A_i}, g_A\}\|_{L^{p'}(\mathbb{R}^n)}\\ &\leqslant \sup_{A\in\mathscr{F}} \|g_A\|_{L^{p'}(\mathbb{R}^n)}.\end{aligned}$$

By the monotone convergence theorem, we have

$$\|g\|_{L^{p'}(\mathbb{R}^n)} = \lim_{i\to\infty} \|\max\{g_{A_i}, g_A\}\|_{L^{p'}(\mathbb{R}^n)} = \|\max\{g, g_A\}\|_{L^{p'}(\mathbb{R}^n)},$$

from which the claim follows. ∎

(13) Assume that $f \in L^p(\mathbb{R}^n)$. Let s be a simple function that is zero outside a set of finite measure. Then $s \in L^p(\mathbb{R}^n)$ (exercise). By (8), we have

$$L(s) = \int_{\mathbb{R}^n} s g_{\{s\neq 0\}}\, d\mu = \int_{\mathbb{R}^n} s g\, d\mu.$$

Since μ is σ-finite, there exists a sequence (s_i) of simple functions s_i, $i=1,2,\dots$, such that every s_i is zero outside a set of finite measure (depending on i), $|s_i(x)| \leqslant |f(x)|$ for μ-almost every $x \in \mathbb{R}^n$ and

$$f(x) = \lim_{i\to\infty} s_i(x)$$

for μ-almost every $x \in \mathbb{R}^n$ (exercise). By the dominated convergence theorem, we have $s_i \to f$ in $L^p(\mathbb{R}^n)$ as $i \to \infty$. By the continuity of L and the dominated convergence theorem, we have

$$\begin{aligned} L(f) &= \lim_{i\to\infty} L(s_i) = \lim_{i\to\infty} \int_{\mathbb{R}^n} s_i g\,d\mu \\ &= \int_{\mathbb{R}^n} (\lim_{i\to\infty} s_i) g\,d\mu = \int_{\mathbb{R}^n} f g\,d\mu \end{aligned}$$

for every $f \in L^p(\mathbb{R}^n)$. The proof of $\|L\| = \|g\|_{p'}$ is the same as in (7) and (9).

To show the uniqueness of g, assume that there exist $g_1, g_2 \in L^{p'}(\mathbb{R}^n)$ such that

$$L(f) = \int_{\mathbb{R}^n} f g_1\,d\mu \quad \text{and} \quad L(f) = \int_{\mathbb{R}^n} f g_2\,d\mu$$

for every $f \in L^p(\mathbb{R}^n)$. As in (8) it follows that

$$\int_{\mathbb{R}^n} f(g_1 - g_2)\,d\mu = 0$$

for every $f \in L^p(\mathbb{R}^n)$. Let $A_i \subset A_{i+1}$ be μ-measurable sets with $\mu(A_i) < \infty$ and $\mathbb{R}^n = \bigcup_{i=1}^\infty A_i$. For example, we may choose $A_i = B(0,i)$, $i=1,2,\dots$. Since $\mu(A_i) < \infty$, we have $f = \chi_{A_i} \operatorname{sign}(g_1 - g_2) \in L^p(A_i)$ and

$$\int_{A_i} |g_1 - g_2|\,d\mu = \int_{\mathbb{R}^n} \chi_{A_i} \operatorname{sign}(g_1 - g_2)(g_1 - g_2)\,d\mu = \int_{\mathbb{R}^n} f(g_1 - g_2)\,d\mu = 0$$

for every $i = 1,2,\dots$. This implies that $g_1(x) = g_2(x)$ for μ-almost every $x \in A_i$. Since this holds for every $i = 1,2,\dots$, we conclude that $g_1(x) = g_2(x)$ for μ-almost every $x \in \mathbb{R}^n$.

(14) Finally we remove the assumption that L is a positive functional by showing that every bounded linear functional $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ can be represented as a difference of two bounded positive functionals. Let $L : L^p(\mathbb{R}^n) \to \mathbb{R}$ be a bounded linear functional. For $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$, let

$$L^+(f) = \sup_{0 \leqslant g \leqslant f} L(g) \quad \text{and} \quad L^-(f) = -\inf_{0 \leqslant g \leqslant f} L(g).$$

CLAIM: $L^+(f_1 + f_2) = L^+(f_1) + L^+(f_2)$ for every $f_1, f_2 \in L^p(\mathbb{R}^n)$ with $f_1 \geqslant 0$ and $f_2 \geqslant 0$.

Reason. Let $g_i, f_i \in L^p(\mathbb{R}^n)$ with $0 \leqslant g_i \leqslant f_i$, $i = 1,2$. Then

$$L(g_1) + L(g_2) = L(g_1 + g_2) \leqslant L^+(f_1 + f_2).$$

By taking supremums over g_1 and g_2, we get

$$L^+(f_1)+L^+(f_2) \leqslant L^+(f_1+f_2).$$

To prove the reverse inequality, let $g \in L^p(\mathbb{R}^n)$ with $0 \leqslant g \leqslant f_1+f_2$ and let $g_1 = \min\{g, f_1\}$ and $g_2 = g - g_1$. Then $0 \leqslant g_i \leqslant f_i$, $i = 1,2$, and we have

$$L(g) = L(g_1+g_2) = L(g_1)+L(g_2) \leqslant L^+(f_1)+L^+(f_2).$$

By taking supremum over g, we get

$$L^+(f_1+f_2) \leqslant L^+(f_1)+L^+(f_2).$$

It follows that $L^+(f_1+f_2) = L^+(f_1)+L^+(f_2)$. ■

CLAIM: $L^+(af) = aL^+(f)$ for every $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$ and $a \geqslant 0$.

Reason.

$$L^+(af) = \sup_{0\leqslant g\leqslant af} L(g) = \sup_{0\leqslant h\leqslant f} L(ah) = a \sup_{0\leqslant h\leqslant f} L(h) = aL^+(f).$$

■

Moroever, for $g \in L^p(\mathbb{R}^n)$ with $0 \leqslant g \leqslant f$, we have

$$L(g) \leqslant \|L\| \|g\|_p \leqslant \|L\| \|f\|_p.$$

By taking supremum over g, we get

$$L^+(f) \leqslant \|L\| \|f\|_p$$

for every $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$. A similar argument shows that

$$L^-(f) \leqslant \|L\| \|f\|_p$$

for every $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$ (exercise).

Let $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$. Then

$$\begin{aligned} L^+(f) - L(f) &= \sup_{0\leqslant g\leqslant f} L(g) - L(f) = \sup_{0\leqslant g\leqslant f} (L(g)-L(f)) \\ &= \sup_{0\leqslant g\leqslant f} -(L(f)-L(g)) = \sup_{0\leqslant g\leqslant f} -L(f-g) \\ &= -\inf_{0\leqslant f-g\leqslant f} L(f-g) = L^-(f), \end{aligned}$$

which shows that $L(f) = L^+(f) - L^-(f)$ for every $f \in L^p(\mathbb{R}^n)$, $f \geqslant 0$.

Define operators $L^+ : L^p(\mathbb{R}^n) \to \mathbb{R}$ and $L^- : L^p(\mathbb{R}^n) \to \mathbb{R}$ by

$$L^+(f) = L^+(f^+) - L^+(f^-) \quad \text{and} \quad L^-(f) = L^-(f^+) - L^-(f^-).$$

The operators L^+ and L^- are linear, bounded and positive. In addition, we have $L(f) = L^+(f) - L^-(f)$ for every $f \in L^p(\mathbb{R}^n)$.

By (13), there exist unique nonnegative $g_1, g_2 \in L^{p'}(\mathbb{R}^n)$ such that

$$L^+(f) = \int_{\mathbb{R}^n} f g_1 \, d\mu \quad \text{and} \quad L^-(f) = \int_{\mathbb{R}^n} f g_2 \, d\mu$$

for every $f \in L^p(\mathbb{R}^n)$. The function $g = g_1 - g_2$ satisfies

$$L(f) = \int_{\mathbb{R}^n} f g \, d\mu$$

for every $f \in L^p(\mathbb{R}^n)$. The fact that $\|L\| = \|g\|_{p'}$ follows as in Theorem 5.1. □

Remarks 5.4:

(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 the inner product.

(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 5.5:

(1) It holds that (exercise)

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}^n} f g \, dx : g \in L^{p'}(\mathbb{R}^n), \|g\|_{p'} \leqslant 1\right\}, \quad 1 \leqslant p < \infty.$$

(2) Since $C_0(\mathbb{R}^n)$ is dense in $L^{p'}(\mathbb{R}^n)$ for $1 < p < \infty$, we have

$$\|f\|_p = \sup\left\{\int_{\mathbb{R}^n} f g \, dx : g \in C_0(\mathbb{R}^n), \|g\|_{p'} \leqslant 1\right\}.$$

(3) To show the uniqueness of g, assume that there exist $g_1, g_2 \in L^{p'}(\mathbb{R}^n)$ such that

$$L(f) = \int_{\mathbb{R}^n} f g_1 \, d\mu \quad \text{and} \quad L(f) = \int_{\mathbb{R}^n} f g_2 \, d\mu$$

for every $f \in L^p(\mathbb{R}^n)$. As in (8) and (13) in the proof of Theorem 5.2 it follows that

$$\int_{\mathbb{R}^n} f(g_1 - g_2) \, d\mu = 0$$

for every $f \in L^p(\mathbb{R}^n)$. By choosing $f = \frac{\chi_{B(x,r)}}{\mu(B(x,r))}$, where $x \in \mathbb{R}^n$ and $r > 0$, we have

$$\frac{1}{\mu(B(x,r))} \int_{B(x,r)} (g_1 - g_2) \, d\mu = 0$$

for every $x \in \mathbb{R}^n$ and $r > 0$. By the Lebesgue differentiation theorem (Theorem 4.33), we conclude that $g_1(x) - g_2(x) = 0$ for μ-almost every $x \in \mathbb{R}^n$. This gives an alternative way to prove the uniqueness in Theorem 5.2.

5.2 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 $n,m=1,2,\dots$. 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 $f:\mathbb{R}^n\to\mathbb{R}^m$. The space $C_0(\mathbb{R}^n;\mathbb{R}^m)$ is relevant since it is dense in many function spaces. For example, $C_0(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1\leqslant p<\infty$, see Theorem 1.57.

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 $x\in\mathbb{R}^n$. Let $L:C_0(\mathbb{R}^n;\mathbb{R}^m)\to\mathbb{R}$,

$$L(f)=\int_{\mathbb{R}^n} f\cdot\sigma\,d\mu$$

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

$$\begin{aligned}|L(f)|&=\left|\int_{\mathbb{R}^n} f\cdot\sigma\,d\mu\right|\leqslant\int_{\mathbb{R}^n}|f\cdot\sigma|\,d\mu\leqslant\int_{\mathbb{R}^n}|f||\sigma|\,d\mu\\&=\int_{\mathbb{R}^n}|f|\,d\mu\leqslant\mu(\operatorname{supp} f)\leqslant\mu(K)<\infty,\end{aligned}$$

which implies

$$\sup\left\{|L(f)| : f\in C_0(\mathbb{R}^n;\mathbb{R}^m),\ |f|\leqslant 1,\ \operatorname{supp} f\subset K\right\}<\infty.$$

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 5.6 (Riesz representation theorem). Assume that $L:C_0(\mathbb{R}^n;\mathbb{R}^m)\to\mathbb{R}$ is a linear functional satisfying

$$\sup\left\{L(f) : f\in C_0(\mathbb{R}^n;\mathbb{R}^m),\ |f|\leqslant 1,\ \operatorname{supp} f\subset K\right\}<\infty \tag{5.7}$$

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 $f \in C_0(\mathbb{R}^n; \mathbb{R}^m)$.

T H E M O R A L : 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. This gives a method to construct Radon measures and motivates the study of Radon measures. The role of σ is just to assign a sign so that the measure μ is nonnegative.

Example 5.8. 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.

Now we are ready for the proof of Theorem 5.6.

Proof. (1) For an open set $U \subset \mathbb{R}^n$, we define a variation measure μ 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 an arbitrary $A \subset \mathbb{R}^n$, we set

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

(2) C L A I M : μ is an outer measure.

Reason. Let U and U_i, $i = 1, 2, \dots$, be open subsets of $\mathbb{R}^n$ such that $U \subset \bigcup_{i=1}^{\infty} U_i$. Let $f \in C_0(\mathbb{R}^n; \mathbb{R}^m)$ 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 exist 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 (Theorem 3.29) 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

$$\sum_{i=1}^{k} \varphi_i(x) = 1 \quad \text{for every} \quad x \in \operatorname{supp} f.$$

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

$$|L(f)|=\left|L\left(\sum_{i=1}^{k}f\varphi_i\right)\right|=\left|\sum_{i=1}^{k}L(f\varphi_i)\right|\leqslant\sum_{i=1}^{k}|L(f\varphi_i)|\leqslant\sum_{i=1}^{\infty}\mu(U_i).$$

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 $i=1,2,\dots$, let U_i be an open set 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) CLAIM: 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

$$U_i=\left\{x\in\mathbb{R}^n:\operatorname{dist}(x,A_i)<\tfrac{1}{3}\operatorname{dist}(A_1,A_2)\right\},\quad i=1,2.$$

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

$$\mu(U_i) < \mu(A) + \tfrac{1}{i}, \quad i = 1, 2, \dots.$$

Thus

$$\mu(A) \leqslant \mu\left(\bigcap_{i=1}^{\infty} U_i\right) \leqslant \mu(U_i) < \mu(A) + \tfrac{1}{i}$$

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 (5.7) 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 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 $a \geqslant 0$ and $f \in C_0^+(\mathbb{R}^n)$.

(5) C L A I M : $\nu(f_1 + f_2) = \nu(f_1) + \nu(f_2)$ 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) CLAIM: $\nu(f) = \int_{\mathbb{R}^n} f\,d\mu$ 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

$$\mu(U_i) - \frac{\varepsilon}{k} \leqslant \nu(h_i) \leqslant \mu(U_i), \quad i = 1,\dots,k.$$

Let

$$A = \left\{x \in \mathbb{R}^n : f(x)\left(1 - \sum_{i=1}^k h_i(x)\right) > 0\right\}.$$

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

$$\begin{aligned}\nu(f) &= \nu\left(f - f\sum_{i=1}^{k} h_i\right) + \nu\left(f\sum_{i=1}^{k} h_i\right) \\ &\leq \varepsilon\|f\|_\infty + \sum_{i=1}^{k} \nu(f h_i) \\ &\leq \varepsilon\|f\|_\infty + \sum_{i=1}^{k} t_i \mu(U_i)\end{aligned}$$

and

$$\nu(f) \geq \sum_{i=1}^{k} \nu(f h_i) \geq \sum_{i=1}^{k} t_{i-1}\left(\mu(U_i) - \frac{\varepsilon}{k}\right) \geq \sum_{i=1}^{k} t_{i-1}\mu(U_i) - t_k\varepsilon.$$

Since

$$\sum_{i=1}^{k} t_{i-1}\mu(U_i) \leq \int_{\mathbb{R}^n} f\,d\mu \leq \sum_{i=1}^{k} t_i \mu(U_i),$$

we have

$$\begin{aligned}\left|\nu(f) - \int_{\mathbb{R}^n} f\,d\mu\right| &\leq \sum_{i=1}^{k} (t_i - t_{i-1})\mu(U_i) + \varepsilon\|f\|_\infty + \varepsilon t_k \\ &\leq \varepsilon\mu(\operatorname{supp} f) + 3\varepsilon\|f\|_\infty.\end{aligned}$$

■

(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

$$\begin{aligned}|\nu_e(f)| = |L(fe)| &\leq \sup\{|L(g)| : g \in C_0(\mathbb{R}^n;\mathbb{R}^m), |g| \leq |f|\} \\ &= \nu(|f|) = \int_{\mathbb{R}^n} |f|\,d\mu.\end{aligned}$$

Thus we can extend ν_e to a bounded linear functional on $L^1(\mathbb{R}^n;\mu)$. By Theorem 5.2 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 $e_1, \dots, e_m$ 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

$$L(f) = \sum_{i=1}^{m} L((f\cdot e_i)e_i) = \sum_{i=1}^{m} \int_{\mathbb{R}^n} (f\cdot e_i)\sigma_{e_i}\,d\mu = \int_{\mathbb{R}^n} f\cdot\sigma\,d\mu.$$

■

(8) CLAIM: $|\sigma(x)| = 1$ 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

$$\mu(U) \leqslant \int_U |\sigma| \, d\mu.$$

Thus

$$\mu(U) = \int_U |\sigma| \, d\mu \quad \text{for every open} \quad U \subset \mathbb{R}^n.$$

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 5.9. 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 unique 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 5.10:

(1) If $f, g \in C_0(\mathbb{R}^n)$ and $f \geqslant g$, then $L(f) - L(g) = L(f - g) \geqslant 0$ and thus $L(f) \geqslant L(g)$.

(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 let $\varphi \in C_0^\infty(\mathbb{R}^n)$ be such that $\varphi = 1$ on K and $0 \leqslant \varphi \leqslant 1$. 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 that

$$L(f) \leqslant c\|f\|_\infty$$

with $c = L(\varphi)$. The mapping L can be extended to a linear mapping from $C_0(\mathbb{R}^n)$ to $\mathbb{R}$, which satisfies the assumptions in the Riesz representation theorem, see Theorem 5.6 (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).

To prove the uniqueness of μ, assume that there exist two measures μ_1 and μ_2 such that

$$L(f) = \int_{\mathbb{R}^n} f\, d\mu_1 \quad \text{and} \quad L(f) = \int_{\mathbb{R}^n} f\, d\mu_2$$

for every $f \in C_0(\mathbb{R}^n)$. Since μ is a Radon measure, it is enough to show that $\mu_1(K) = \mu_2(K)$ for every compact set $K \subset \mathbb{R}^n$. Let $K \subset \mathbb{R}^n$ be a compact set and let $\varepsilon > 0$. By the properties of Radon measure, there exists an open set $U \supset K$ with $\mu(U) \leqslant \mu(K) + \varepsilon$. Assume that $U \subset \mathbb{R}^n$ is an open set and that $F \subset G$ a compact set. As in the proof of Theorem 1.57, there exists a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ such that $0 \leqslant f(x) \leqslant 1$ for every $x \in \mathbb{R}^n$, $f(x) = 1$ for every $x \in K$ and $\operatorname{supp} f$ is a compact subset of U. This implies

$$\begin{aligned}\mu_1(K) &= \int_{\mathbb{R}^n} \chi_K\, d\mu_1 \leqslant \int_{\mathbb{R}^n} f\, d\mu_1 = L(f) = \int_{\mathbb{R}^n} f\, d\mu_2 \\ &\leqslant \int_{\mathbb{R}^n} \chi_U\, d\mu_2 = \mu_2(U) \leqslant \mu(K) + \varepsilon.\end{aligned}$$

This shows that $\mu_1(K) \leqslant \mu_2(K)$. The reverse inequality holds by switching the roles of μ_1 and μ_2. □

Remark 5.11. The Riesz representation theorem holds in a much more general context. The underlying space can be any locally compact Hausdorff space X instead of $\mathbb{R}^n$.

5.3 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

$$\lim_{i\to\infty} x_i^*(x) = x^*(x) \quad \text{for every} \quad x \in X.$$

Both weak and weak star convergences are pointwise convergences tested on every element of the space X^* and X, respectively.

By the Riesz representation theorem (Theorem 5.6), 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 5.12. The sequence (μ_i) of Radon measures μ_i, $i = 1,2,\dots$, 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 5.13:

(1) Let ϕ_ε, $\varepsilon > 0$, be the standard mollifier in Example 3.10 (2) and let μ_i be a Radon measure on $\mathbb{R}^n$ defined by

$$\mu_i(A) = \mu_{\varepsilon_i}(A) = \int_A \phi_{\varepsilon_i}(y)\,dy,$$

with $\varepsilon_i \to 0$ as $i \to \infty$, for every Borel set $A \subset \mathbb{R}^n$. As in Example 3.10 (2), we have

$$\lim_{i\to\infty}\int_{\mathbb{R}^n} f(y)\,d\mu_i(y) = \lim_{i\to\infty}\int_{\mathbb{R}^n} f(y)\varphi_i(y)\,dy = f(0)$$

for every $f \in C_0(\mathbb{R}^n)$. This implies that $\mu_i \rightharpoonup \delta_0$ as $i \to \infty$, where δ_0 is Dirac's measure at 0.

(2) Let δ_i be Dirac's measure at $i = 1,2,\dots$ on $\mathbb{R}$. Then $\delta_i \rightharpoonup 0$ as $i \to \infty$ (exercise).

(3) Let

$$\mu_i = \frac{1}{i}\left(\delta_{\frac{1}{i}} + \delta_{\frac{2}{i}} + \dots + \delta_{\frac{i}{i}}\right), \quad i = 1,2,\dots.$$

Then for every $f \in C_0(\mathbb{R})$, we have

$$\int_{\mathbb{R}} f\,d\mu_i = \sum_{j=1}^{i} \tfrac{1}{i} f\left(\tfrac{j}{i}\right) \to \int_0^1 f(x)\,dx,$$

since these are Riemann sums in $[0,1]$. Thus $\mu_i \rightharpoonup m^1\lfloor_{[0,1]}$ as $i \to \infty$. Here $\lfloor$ denotes the restriction of a measure to a subset.

Lemma 5.14. 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. (1) Let $K \subset \mathbb{R}^n$ be compact and let U be an open set with $K \subset U$. Let $f \in C_0(\mathbb{R}^n)$ such that $0 \leqslant f \leqslant 1$, $\operatorname{supp} f \subset U$ and $f = 1$ on K. Then

$$\begin{aligned}\mu(U) &= \int_U 1\,d\mu \geqslant \int_U f\,d\mu = \int_{\mathbb{R}^n} f\,d\mu \\ &= \lim_{i\to\infty}\int_{\mathbb{R}^n} f\,d\mu_i \geqslant \limsup_{i\to\infty}\mu_i(K).\end{aligned}$$

Taking infimum over all open sets $U \supset K$, we have

$$\limsup_{i\to\infty}\mu_i(K) \leqslant \inf\{\mu(U) : U \supset K, U \text{ open}\} = \mu(K).$$

(2) Let U be open and $K \subset U$ compact. Let f be the same function as in (1). 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 5.15. Let (μ_i) be a sequence of Radon measures μ_i, $i = 1, 2, \dots$, 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 $\mu_{i_j} \rightharpoonup \mu$ as $j \to \infty$.

THE MORAL: Every locally bounded sequence of Radon measures on $\mathbb{R}^n$ has a weakly converging subsequence.

Proof. (1) Assume first that $M = \sup_i \mu_i(\mathbb{R}^n) < \infty$.

(2) Let $\{f_k\}_{k=1}^\infty$ be a countable dense subset of $C_0(\mathbb{R}^n)$ with respect to $\|\cdot\|_\infty$ norm (exercise). We apply a diagonal argument. The assumption in (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 $\left(\int_{\mathbb{R}^n} f_1\,d\mu_i\right)$ 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

$$\int_{\mathbb{R}^n} f_1\,d\mu_i^1 \xrightarrow{i\to\infty} a_1.$$

Recursively, there exists a subsequence (μ_i^k) of (μ_i^{k-1}) and $a_k \in \mathbb{R}$ such that

$$\int_{\mathbb{R}^n} f_k\,d\mu_i^k \xrightarrow{i\to\infty} a_k.$$

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 a 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 a functional $L:S\to\mathbb{R}$ by setting

$$L(g)=\sum_{k=1}^m \lambda_k a_k, \quad \text{where} \quad g=\sum_{k=1}^m \lambda_k f_k.$$

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 linear functional on S. Moreover,

$$|L(g)|=\lim_{i\to\infty}\left|\int_{\mathbb{R}^n} g\,d\mu_i^i\right|\leqslant \sup_i(\|g\|_\infty\mu_i^i(\mathbb{R}^n))\leqslant M\|g\|_\infty. \tag{5.16}$$

This shows that L is a bounded functional on S.

(3) The functional $L:S\to\mathbb{R}$ 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 $\{f_k\}_{j=1}^\infty$ is dense in $C_0(\mathbb{R}^n)$, there exists a sequence (f_j) such that $\|f_j-f\|_\infty\to 0$ as $j\to\infty$. It follows from (5.16) that $(L(f_j))$ is a Cauchy sequence in $\mathbb{R}$ and thus it converges. Let

$$L(f)=\lim_{j\to\infty}L(f_j).$$

It follows from (5.16) that L is a well-defined functional in $C_0(\mathbb{R}^n)$ and that (5.16) holds for every $f\in C_0(\mathbb{R}^n)$. ∎

(4) CLAIM: $L:C_0(\mathbb{R}^n)\to\mathbb{R}$ is a positive functional.

Reason. If $f\in C_0(\mathbb{R}^n)$ with $f\geqslant 0$, and $\|f_j-f\|_\infty\to 0$ as $j\to\infty$ with $f_j\in S$, then

$$\liminf_{j\to\infty}(\min_{\mathbb{R}^n} f_j)\geqslant 0.$$

We observe that, if $\min_{\mathbb{R}^n} f_j<0$, then

$$\int_{\mathbb{R}^n} f_j\,d\mu_i^i\geqslant \mu_i^i(\mathbb{R}^n)\min_{\mathbb{R}^n} f_j\geqslant \sup_i \mu_i(\mathbb{R}^n)\min_{\mathbb{R}^n} f_j=M\min_{\mathbb{R}^n} f_j.$$

On the other hand, if $\min_{\mathbb{R}^n} f_j \geqslant 0$, then

$$\int_{\mathbb{R}^n} f_j \, d\mu_i^i \geqslant 0.$$

It follows that

$$\int_{\mathbb{R}^n} f_j \, d\mu_i^i \geqslant M \min\{0, \min_{\mathbb{R}^n} f_j\}$$

for every $i, j = 1, 2, \dots$. By (3) we obtain

$$\begin{aligned} L(f) &= \lim_{j\to\infty} L(f_j) = \lim_{j\to\infty} \lim_{i\to\infty} \int_{\mathbb{R}^n} f_j \, d\mu_i^i \\ &\geqslant M \liminf_{j\to\infty} \min\{0, \min_{\mathbb{R}^n} f_j\} \geqslant 0. \end{aligned}$$

■

According to the Riesz representation theorem (Theorem 5.9) there exists a Radon measure μ on $\mathbb{R}^n$ such that

$$L(f) = \int_{\mathbb{R}^n} f \, d\mu \quad \text{for every} \quad f \in C_0(\mathbb{R}^n).$$

(5) CLAIM: $\mu_i^i \rightharpoonup \mu$ as $i \to \infty$.

Reason. Let $\varepsilon > 0$ and $f \in C_0(\mathbb{R}^n)$. Let $g \in S$ be 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 $j = 1, 2, \dots$ there exists a subsequence (μ_i^j) of (μ_i^{j-1}) such that

$$\mu_i^j \lfloor_{B(0,j)} \rightharpoonup \nu^j, \quad \text{as} \quad i \to \infty,$$

where ν^j is a Radon measure with $\nu^j(\mathbb{R}^n \setminus \overline{B}(0,j)) = 0$. Here $\lfloor$ denotes the restriction of a measure to a subset. The diagonal sequence (μ_i^i) satisfies

$$\mu_i^i \lfloor_{B(0,j)} \rightharpoonup \nu^j \quad \text{as} \quad i \to \infty \quad \text{for every} \quad j = 1, 2, \dots.$$

Here $\lfloor$ denotes the restriction of a measure to a subset. 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 $i \to \infty$. □

5.4 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 $1 \leqslant p < \infty$. We only discuss the case when the underlying measure is the Lebesgue measure although similar results hold also for $L^p(\mathbb{R}^n;\mu)$, where μ is a Radon measure.

Definition 5.17. Let $1 \leqslant p \leqslant \infty$. A sequence (f_i) of functions in $L^p(\mathbb{R}^n)$ converges weakly (weak 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 5.18. $f_i \to f$ 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

$$\begin{aligned} \left| \int_{\mathbb{R}^n} f_i g \, dx - \int_{\mathbb{R}^n} f g \, dx \right| &\leqslant \int_{\mathbb{R}^n} |f_i - f||g| \, dx \\ &\leqslant \|f_i - f\|_p \|g\|_{p'} \xrightarrow{i\to\infty} 0 \end{aligned}$$

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 5.19:

(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

$$\|f\|_p < \liminf_{i\to\infty} \|f_i\|_p.$$

THE MORAL: Sequences of rapidly oscillating functions provide examples of weakly converging sequences that do not converge strongly.

(2) (Concentration) Let $f_i : (-1, 1) \to \mathbb{R}$,

$$f_i(x) = \begin{cases} i, & 0 \leqslant x \leqslant \frac{1}{i}, \\ 0, & \text{otherwise.} \end{cases}$$

Then $f_i \to \delta_0$ weakly as measures in $(-1, 1)$. Observe that the sequence (f_i) converges weakly to zero as measures in $(0, 1)$, but (f_i) does not converge weakly in $L^1((0, 1))$.

THE MORAL: Sequences of concentrating functions provide examples of weakly converging sequences that do not converge strongly.

(3) Let $1 < p < \infty$ and $f_i : (-1,1) \to \mathbb{R}$,

$$f_i(x) = \begin{cases} i^{\frac{1}{p}}, & 0 \leqslant x \leqslant \frac{1}{i}, \\ 0, & \text{otherwise.} \end{cases}$$

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,

$$\int_{(-1,1)} |f_i|^q \, dx = i^{\frac{q}{p}-1} \xrightarrow{i\to\infty} 0.$$

In particular, the norms $\|f_i\|_q$, $q < p$, do not concentrate.

The next result shows that any weakly converging sequence is bounded.

Theorem 5.20. If $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$ with $1 \leqslant p \leqslant \infty$, 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. We apply similar arguments as in the proof of Theorem 5.1. If $\|f\|_p = 0$, the claim is clear. Hence we may assume that $\|f\|_p > 0$.

$\boxed{1 < p < \infty}$ The function $g = |f|^{\frac{p}{p'}} \operatorname{sign} f$ belongs $L^{p'}(\mathbb{R}^n)$, since

$$\|g\|_{p'} = \left(\int_{\mathbb{R}^n} |g|^{p'} \, dx\right)^{\frac{1}{p'}} = \left(\int_{\mathbb{R}^n} |f|^p \, dx\right)^{\frac{1}{p'}} = \|f\|_p^{\frac{p}{p'}} < \infty.$$

Since $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$, we have

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} f_i g \, dx = \int_{\mathbb{R}^n} f g \, dx = \int_{\mathbb{R}^n} |f|^{\frac{p}{p'}} \underbrace{f \operatorname{sign} f}_{=|f|} \, dx = \int_{\mathbb{R}^n} |f|^p \, dx = \|f\|_p^p.$$

On the other hand, by Hölder's inequality

$$\left|\int_{\mathbb{R}^n} f_i g \, dx\right| \leqslant \|f_i\|_p \|g\|_{p'} = \|f_i\|_p \|f\|_p^{\frac{p}{p'}}$$

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

$$\|f\|_p^p = \lim_{i\to\infty} \left|\int_{\mathbb{R}^n} f_i g \, dx\right| \leqslant \liminf_{i\to\infty} \|f_i\|_p \|f\|_p^{\frac{p}{p'}}.$$

The claim follows by dividing through by $\|f\|_p^{\frac{p}{p'}} > 0$.

$\boxed{p = 1}$ The function $g = \operatorname{sign} f$ belongs to $L^\infty(\mathbb{R}^n)$ and $\|g\|_\infty \leqslant 1$. Since $f_i \to f$ weakly in $L^1(\mathbb{R}^n)$ as $i \to \infty$, we have

$$\lim_{i\to\infty} \int_{\mathbb{R}^n} f_i g \, dx = \int_{\mathbb{R}^n} f g \, dx = \int_{\mathbb{R}^n} f \operatorname{sign} f \, dx = \int_{\mathbb{R}^n} |f| \, dx = \|f\|_1.$$

On the other hand, by Hölder's inequality

$$\left|\int_{\mathbb{R}^n} f_i g\,dx\right| \leqslant \|f_i\|_1\|g\|_\infty \leqslant \|f_i\|_1$$

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

$$\|f\|_1 = \lim_{i\to\infty}\left|\int_{\mathbb{R}^n} f_i g\,dx\right| \leqslant \liminf_{i\to\infty}\|f_i\|_1.$$

$\boxed{p=\infty}$ Exhaust $\mathbb{R}^n$ with measurable sets $A_j \subset A_{j+1}$, $\mathbb{R}^n = \bigcup_{j=1}^\infty A_j$ with $|A_j| < \infty$ for every $j = 1,2,\ldots$. For example, we may choose $A_j = B(0,j)$ for every $j = 1,2,\ldots$. For $0 < \varepsilon < \|f\|_{L^\infty(\mathbb{R}^n)}$, let

$$A_{\varepsilon,j} = \{x \in A_j : |f(x)| \geqslant \|f\|_\infty - \varepsilon\} \quad \text{and} \quad g_{\varepsilon,j} = \frac{\chi_{A_{\varepsilon,j}}\operatorname{sign} f}{|A_{\varepsilon,j}|}.$$

Let $A_\varepsilon = \{x \in \mathbb{R}^n : |f(x)| \geqslant \|f\|_\infty - \varepsilon\}$. We observe that $A_{\varepsilon,j} \subset A_{\varepsilon,j+1}$, $j = 1,2,\ldots$, and $A_\varepsilon = \bigcup_{j=1}^\infty A_{\varepsilon,j}$. Since $|A_\varepsilon| > 0$ and we have

$$0 < |A_\varepsilon| = \left|\bigcup_{j=1}^\infty A_{\varepsilon,i}\right| = \lim_{j\to\infty}|A_{\varepsilon,j}|$$

and, consequently, there exists j such that $|A_{\varepsilon,j}| > 0$. On the other hand, $|A_{\varepsilon,j}| \leqslant |B(0,j)| < \infty$.

We observe that $g_{\varepsilon,j} \in L^1(\mathbb{R}^n)$, since

$$\|g_{\varepsilon,j}\|_{L^1(\mathbb{R}^n)} = \int_{\mathbb{R}^n}\left|\frac{\chi_{A_{\varepsilon,j}}\operatorname{sign} f}{|A_{\varepsilon,j}|}\right| dx \leqslant \int_{\mathbb{R}^n}\frac{\chi_{A_{\varepsilon,j}}}{|A_{\varepsilon,j}|}\,dx = \frac{|A_{\varepsilon,j}|}{|A_{\varepsilon,j}|} = 1.$$

Since $f_i \to f$ weakly (weak star) in $L^\infty(\mathbb{R}^n)$ as $i \to \infty$, we have

$$\begin{aligned}
\lim_{i\to\infty}\int_{A_{\varepsilon,j}} f_i g_{\varepsilon,j}\,dx &= \lim_{i\to\infty}\int_{\mathbb{R}^n} f_i g_{\varepsilon,j}\,dx \\
&= \int_{\mathbb{R}^n} f g_{\varepsilon,j}\,dx = \int_{A_{\varepsilon,j}} f g_{\varepsilon,j}\,dx \\
&= \int_{A_{\varepsilon,j}} \frac{|f|}{|A_{\varepsilon,j}|}\,dx \geqslant \|f\|_\infty - \varepsilon.
\end{aligned}$$

On the other hand, by Hölder's inequality

$$\left|\int_{A_{\varepsilon,j}} f_i g_{\varepsilon,j}\,dx\right| \leqslant \|f_i\|_\infty\|g_{\varepsilon,j}\|_1 = \|f_i\|_\infty$$

for every $i = 1,2,\ldots$. It follows that

$$\|f\|_\infty - \varepsilon \leqslant \lim_{i\to\infty}\left|\int_{A_{\varepsilon,j}} f_i g_{\varepsilon,j}\,dx\right| \leqslant \liminf_{i\to\infty}\|f_i\|_\infty.$$

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

Remark 5.21. Let $1 < p < \infty$. If $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$, it does not follow that that $\lim_{i\to\infty} \|f_i\|_p = \|f\|_p$. Nor does the reverse implication hold true. Example 5.19 (1) gives a counter example for both claims. The following result explains the situation: If $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ with $1 < p < \infty$ and $\lim_{i\to\infty} \|f_i\|_p = \|f\|_p$, then $f_i \to f$ strongly in $L^p(\mathbb{R}^n)$. This result will not be proved here.

Remark 5.22. 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 5.23. Let $1 < p < \infty$. Assume that the sequence (f_i) of functions $f_i \in L^p(\mathbb{R}^n)$, $i = 1, 2, \dots$, 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 5.24. 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, let μ be 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 $A \subset \mathbb{R}^n$. By Theorem 5.15, 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 5.25. Let μ be the one-dimensional Lebesgue measure on $\mathbb{R}$. Then the sequence $f_i = i\chi_{[0,\frac{1}{i}]}$, $i = 1, 2, \dots$, converges in measure to zero, satisfies $\|f_i\|_1 = 1$ for every $i = 1, 2, \dots$, and thus

$$\sup_i \|f_i\|_1 < \infty.$$

CLAIM: (f_i) does not converge weakly in $L^1(\mathbb{R})$.

Reason. For a contradiction assume that there exists $f \in L^1(\mathbb{R})$ such that $f_i \to f$ weakly in $L^1(\mathbb{R})$ as $i \to \infty$. Then

$$\lim_{i\to\infty} \int_{\mathbb{R}} f_i g\,dx = \int_{\mathbb{R}} f g\,dx \quad \text{for every} \quad g \in L^\infty(\mathbb{R}).$$

Let $y \neq 0$ and $g = \frac{1}{2\varepsilon}\chi_{[y-\varepsilon,y+\varepsilon]}$ with $0 < \varepsilon < |y|$. Then

$$0 = \lim_{i\to\infty} \int_{\mathbb{R}} f_i g\,dx = \int_{\mathbb{R}} f g\,dx = \frac{1}{2\varepsilon} \int_{[y-\varepsilon,y+\varepsilon]} f\,dx.$$

By the Lebesgue differentiation theorem (Theorem 4.33), we conclude that

$$0 = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_{[y-\varepsilon,y+\varepsilon]} f\,dx = f(y)$$

for almost every $y \in \mathbb{R}$, $y \neq 0$. Thus $f = 0$ almost everywhere in $\mathbb{R}$ and

$$\lim_{i\to\infty} \int_{\mathbb{R}} f_i g\,dx = \int_{\mathbb{R}} f g\,dx = 0 \quad \text{for every} \quad g \in L^\infty(\mathbb{R}).$$

By letting $g = 1$, we have

$$\int_{\mathbb{R}} f_i g\,dx = \int_{\mathbb{R}} f_i\,dx = 1$$

for every $i = 1, 2, \ldots$. This is a contradiction. ■

However, consider the measures

$$\mu_i(A) = \int_{\mathbb{R}} f_i\,dx = i \int_{A\cap[0,\frac{1}{i}]} 1\,dx = i\left|A \cap \left[0, \tfrac{1}{i}\right]\right|$$

for every Borel set $A \subset \mathbb{R}$.

CLAIM: $\mu_i \rightharpoonup \delta_0$ as $i \to \infty$.

Reason.

$$\begin{aligned}\lim_{i\to\infty} \int_{\mathbb{R}} g\,d\mu_i &= \lim_{i\to\infty} \int_{\mathbb{R}} g f_i\,dx = \lim_{i\to\infty} \frac{1}{i} \int_{[0,\frac{1}{i}]} g\,dx \\ &= g(0) = \int_{\mathbb{R}} g\,d\delta_0\end{aligned}$$

for every $g \in C_0(\mathbb{R})$. ■

Proof. (1) We may assume that $f_i \geqslant 0$ almost everywhere for every $i = 1, 2, \ldots$, for otherwise we may consider the positive and negative parts f_i^+ and f_i^-. (Exercise)

(2) Define Radon measures μ_i, $i = 1, 2, \ldots$, by setting

$$\mu_i(A) = \int_A f_i\,dx \tag{5.26}$$

where $A \subset \mathbb{R}^n$ is a Borel set. Then for each compact set $K \subset \mathbb{R}^n$,

$$\mu_i(K) = \int_K f_i \, dx \leqslant \left(\int_K f_i^p \, dx \right)^{\frac{1}{p}} |K|^{1-\frac{1}{p}}$$

for every $i = 1, 2, \ldots$. This implies $\sup_i \mu_i(K) < \infty$. Thus we may apply Theorem 5.15 to find a Radon measure μ on $\mathbb{R}^n$ and a subsequence (μ_{i_j}), $j = 1, 2, \ldots$, 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 choose an open and bounded set $U \supset A$ such that $|U| < \varepsilon$. By Lemma 5.14 and (5.26), we have

$$\begin{aligned} \mu(U) &\leqslant \liminf_{i\to\infty} \mu_{i_j}(U) = \liminf_{j\to\infty} \int_U f_{i_j} \, dx \\ &\leqslant \liminf_{j\to\infty} \left(\int_U f_{i_j}^p \, dx \right)^{\frac{1}{p}} |U|^{1-\frac{1}{p}} \leqslant c\varepsilon^{1-\frac{1}{p}}. \end{aligned}$$

Thus $\mu(A) = 0$. ■

(4) By the Radon-Nikodym theorem (Theorem 4.23), there exists $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ satisfying

$$\mu(A) = \int_A f \, dx \tag{5.27}$$

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

(5) CLAIM: $f \in L^p(\mathbb{R}^n)$.

Reason. Let $\varphi \in C_0(\mathbb{R}^n)$. By (5.27) and (5.26) we have

$$\begin{aligned} \int_{\mathbb{R}^n} \varphi f \, dx &= \int_{\mathbb{R}^n} \varphi \, d\mu = \lim_{j\to\infty} \int_{\mathbb{R}^n} \varphi \, d\mu_{i_j} \\ &= \lim_{j\to\infty} \int_{\mathbb{R}^n} \varphi f_{i_j} \, dx \\ &\leqslant \sup_i \|f_i\|_p \|\varphi\|_q, \end{aligned}$$

where $\sup_i \|f_i\|_p < \infty$ by the assuption. 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) CLAIM: $f_{i_j} \to f$ 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 let $\varphi \in C_0(\mathbb{R}^n)$ with $\|g-\varphi\|_{p'} < \varepsilon$. Then

$$\begin{aligned}
&\left|\int_{\mathbb{R}^n} f_{i_j} g\,dx - \int_{\mathbb{R}^n} f g\,dx\right| \\
&\leqslant \left|\int_{\mathbb{R}^n} f_{i_j} g\,dx - \int_{\mathbb{R}^n} \varphi f_{i_j}\,dx\right| + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| \\
&\quad + \left|\int_{\mathbb{R}^n} \varphi f\,dx - \int_{\mathbb{R}^n} g f\,dx\right| \\
&\leqslant \|f_{i_j}\|_p \|g-\varphi\|_{p'} + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| + \|f\|_p \|g-\varphi\|_{p'} \\
&\leqslant \varepsilon \sup_i \|f_i\|_p + \left|\int_{\mathbb{R}^n} \varphi f_{i_j}\,dx - \int_{\mathbb{R}^n} \varphi f\,dx\right| + \varepsilon \|f\|_p.
\end{aligned}$$

This implies that

$$\lim_{j\to\infty} \int_{\mathbb{R}^n} f_{i_j} g\,dx = \int_{\mathbb{R}^n} f g\,dx.$$

■

Remark 5.28. 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 5.29. Let $1 < p \leqslant \infty$. Assume that the sequence (f_i) of functions $f_i \in L^p(\mathbb{R}^n)$, $i = 1, 2, \dots$. 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) $\displaystyle\lim_{i\to\infty} \int_A f_i\,dx = \int_A f\,dx$ 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. □

THE MORAL: Property (2) asserts that the averages of the functions f_i converge to the average of f over A.

Recall that weak convergence does not imply strong converge. We close this section by a result which is sometimes useful.

Theorem 5.30. Let $A \subset \mathbb{R}^n$ be a measurable set with $|A| < \infty$ and let $1 < p < \infty$. Assume that $f_i \to f$ weakly in $L^p(A)$ and $f_i \to f$ almost everywhere in A. Then $f_i \to f$ in $L^q(A)$ whenever $1 < q < p$.

THE MORAL: Pointwise convergence and weak convergence in L^p imply strong convergence in L^q, with $q < p$, on bounded sets.

Proof. Since a weakly converging sequence is bounded, we have

$$M = \sup_i \|f_i\|_{L^p(A)} < \infty$$

and

$$\|f\|_{L^p(A)} \leqslant \liminf_{i\to\infty} \|f_i\|_{L^p(A)} \leqslant M.$$

Cavalieri's principle implies that

$$\begin{aligned}\int_A |f_i - f|^q\,dx &= \int_{\{|f_i-f|\leqslant k\}\cap A} |f_i - f|^q\,dx + \int_{\{|f_i-f|>k\}\cap A} |f_i - f|^q\,dx\\ &= \int_{\{|f_i-f|\leqslant k\}\cap A} |f_i - f|^q\,dx\\ &\quad + q\int_k^\infty \lambda^{q-1}|\{|f_i - f| > \lambda\}\cap A|\,d\lambda + k^q|\{|f_i - f| > k\}\cap A|.\end{aligned}$$

By Chebyshev's inequality, we have

$$\begin{aligned}|\{|f_i - f| > \lambda\}\cap A| &\leqslant \frac{1}{\lambda^p}\int_A |f_i - f|^p\,dx \leqslant \frac{1}{\lambda^p}\int_A (|f_i| + |f|)^p\,dx\\ &\leqslant \frac{2^p}{\lambda^p}\int_A (|f_i|^p + |f|^p)\,dx \leqslant \frac{2^{p+1}M^p}{\lambda^p}\end{aligned}$$

for every $i = 1,2,\dots$. It follows that

$$q\int_k^\infty \lambda^{q-1}|\{|f_i - f| > \lambda\}\cap A|\,d\lambda \leqslant \frac{2^{p+1}M^p}{p-q}k^{q-p}$$

and

$$k^q|\{|f_i - f| > k\}\cap A| \leqslant 2^{p+1}M^p k^{q-p}.$$

Let $\varepsilon > 0$. By choosing k large enough so that

$$\max\left\{\frac{2^{p+1}M^p}{p-q}k^{q-p}, 2^{p+1}M^p k^{q-p}\right\} = 2^{p+1}M^p k^{q-p}\max\left\{\frac{1}{p-q}, 1\right\} < \frac{\varepsilon}{2},$$

we arrive at

$$\int_A |f_i - f|^q\,dx \leqslant \int_{\{|f_i-f|\leqslant k\}\cap A} |f_i - f|^q\,dx + \varepsilon$$

for every $i = 1,2,\dots$. Since $\chi_{\{|f_i-f|\leqslant k\}\cap A}|f_i - f|^q \leqslant k^q$ for every $i = 1,2,\dots$, $|A| < \infty$ and $f_i \to f$ in A as $i \to \infty$, the dominated convergence theorem implies that

$$\int_{\{|f_i-f|\leqslant k\}\cap A} |f_i - f|^q\,dx = \int_A \chi_{\{|f_i-f|\leqslant k\}\cap A}|f_i - f|^q\,dx \xrightarrow{i\to\infty} 0.$$

Thus we may choose i_ε large enough so that

$$\int_{\{|f_i-f|\leqslant k\}\cap A} |f_i - f|^q\,dx < \varepsilon$$

for every $i \geqslant i_\varepsilon$. It follows that

$$\int_A |f_i - f|^q\,dx < 2\varepsilon$$

for every $i \geqslant i_\varepsilon$. This shows that $f_i \to f$ in $L^q(A)$ as $i \to \infty$. □

5.5 Mazur's lemma

This section discusses a method to upgrade weak convergence to strong convergence. We are mainly interested in the case of $L^p(\mathbb{R}^n)$ with $1 < p < \infty$, compare to Theorem 5.23, but we consider a general normed space X.

We recall some facts related to the Hanh-Banach theorem, which are needed in the argument. A function $p : X \to \mathbb{R}$ is sublinear, if $p(x+y) \leqslant p(x)+p(y)$ for every $x, y \in X$ and $p(\lambda x) = \lambda p(x)$ for every $x \in X$ and $\lambda \geqslant 0$. In particular, every seminorm on X determines a sublinear function. More generally, if C is a convex, open neighborhood of 0 in a normed space X , then $p : X \to \mathbb{R}$,

$$p(x) = \inf\{\lambda > 0 : \lambda^{-1} x \in C_\varepsilon\},$$

is a sublinear function, called the Minkowski functional associated with C. Let $p : X \to \mathbb{R}$ is sublinear function, let Y be a vector subspace of X and let $f : Y \to \mathbb{R}$ be a linear function with $f(x) \leqslant p(x)$ for every $x \in Y$. The Hahn-Banach theorem asserts that there exist a linear function $\overline{f} : X \to \mathbb{R}$ such that $\overline{f}|_Y = f$, that is $\overline{f}(x) = f(x)$ for every $x \in Y$, and $\overline{f}(x) \leqslant p(x)$ for every $x \in X$. The mapping $\overline{f}$ is called a linear extension of f to X.

Theorem 5.31 (Mazur's lemma). Assume that X is a normed space and that $x_i \to x$ weakly in X as $i \to \infty$. Then for every $\varepsilon > 0$, there exists $m \in \mathbb{N}$ and a convex combination $\sum_{i=1}^m a_i x_i$ with $a_i \geqslant 0$ and $\sum_{i=1}^m a_i = 1$, such that

$$\left\| x - \sum_{i=1}^m a_i x_i \right\| < \varepsilon.$$

If $x_i \to x$ weakly in X, Mazur's lemma gives the existence of a sequence (x_k) of convex combinations

$$\widetilde{x}_i = \sum_{j=1}^{m_i} a_{i,j} x_j, \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{m_i} a_{i,j} = 1,$$

such that $\widetilde{x}_i \to x$ in the norm of X as $i \to \infty$.

THE MORAL: For every weakly converging sequence, there is a sequence of convex combinations that converges strongly. Thus weak convergence is upgraded to strong convergence for a sequence of convex combinations. Observe that some of the coefficients a_i may be zero so that the convex combination is essentially for a subsequence.

Proof. Let C be the set of all convex combinations of x_i, $i = 1, 2, \dots$, that is

$$C = \left\{ \sum_{i=1}^m a_i x_i : a_i \geqslant 0, \sum_{i=1}^m a_i = 1, m \in \mathbb{N} \right\}.$$

By replacing the sequence $(x_i)_{i=1}^{\infty}$ by a sequence $(x_i - x_1)_{i=1}^{\infty}$ and x by $x - x_1$, we may assume that $0 \in C$. For a contradiction, assume that there exists $\varepsilon > 0$ with $\|x - y\| \geqslant 2\varepsilon$ for every $y \in C$. In particular, this implies $x \neq 0$.

The ε-neighbourhood of C defined by

$$C_\varepsilon = \{y \in X : \operatorname{dist}(y, C) < \varepsilon\}$$

is a convex set and an open neighbourhood of 0. Consider the Minkowski functional $p : X \to \mathbb{R}$ associated with C_ε defined by

$$p(y) = \inf\{\lambda > 0 : \lambda^{-1} y \in C_\varepsilon\}.$$

Since C_ε is convex, p is a sublinear function on X. Since C_ε is an open neighbourhood of 0, p is a continuous function on X (exercise).

Let $z \in C_\varepsilon$ and $y \in C$ with $\|z - y\| < \varepsilon$. Since $\|x - y\| \geqslant 2\varepsilon$ for every $y \in C$, by the triangle inequality, we conclude that

$$\|x - z\| \geqslant \|x - y\| - \|y - z\| > 2\varepsilon - \varepsilon = \varepsilon$$

for every $z \in C_\varepsilon$. This implies that there exists $y_0 \in C_\varepsilon$ such that $x = \lambda^{-1} y_0$ with $0 < \lambda < 1$ and $p(y_0) = 1$ and thus

$$p(x) = p(\lambda^{-1} y_0) = \lambda^{-1} p(y_0) = \lambda^{-1} > 1.$$

Consider the vector subspace $Y = \{t y_0 : t \in \mathbb{R}\}$ of X and the linear function $f : Y \to \mathbb{R}$ defined by $f(t y_0) = t$. Then $f(y) \leqslant p(y)$ for every $y \in Y$ and by the Hahn-Banach theorem, there exists an extension of f to a linear functional $\overline{f} : X \to \mathbb{R}$ with $\overline{f}(y) \leqslant p(y)$ for every $y \in X$. Since p is continuous, $\overline{f}$ is continuous and thus $\overline{f} \in X^*$. Since $x_I \to x$ weakly in X, we have

$$\overline{f}(x) = \lim_{i \to \infty} \overline{f}(x_i).$$

Since $x \in Y$ we have $\overline{f}(x) = p(x)$ and since $x_i \in C$ we have $p(x_i) \leqslant 1$ for every $i = 1, 2, \ldots$. It follows that

$$1 < p(x) = \overline{f}(x) = \lim_{i \to \infty} \overline{f}(x_i) \leqslant \liminf_{i \to \infty} p(x_i) \leqslant 1$$

which is a contradiction. □

The following tail version of Mazur's lemma will be useful in applications. Observe that the indexing of the sequence of convex combinations starts from i instead of one.

Lemma 5.32. Assume that X is a normed space and that $x_i \to x$ weakly in X as $i \to \infty$. Then there exists a sequence of convex combinations

$$\widetilde{x}_i = \sum_{j=i}^{m_i} a_{i,j} x_j, \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=i}^{m_i} a_{i,j} = 1,$$

such that $\widetilde{x}_i \to x$ in the norm of X as $i \to \infty$.

Proof. By applying Mazur's lemma repeatedly to the sequences $(x_j)_{j=i}^{\infty}$, $i = 1,2,\dots$, we obtain convex combinations

$$\widetilde{x}_i = \sum_{j=i}^{m_i} a_{i,j} x_j, \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=i}^{m_i} a_{i,j} = 1,$$

such that $\|\widetilde{x}_i - x\| < \frac{1}{i}$. This implies that $\widetilde{x}_i \to x$ in the norm of X as $i \to \infty$. □

We discuss briefly applications of Mazur's lemma for $L^p(\mathbb{R}^n)$ with $1 < p < \infty$. The following process upgrades the mode of convergence stepwise. This is applied, for example, in the direct methods of the calculus of variations.

(1) Assume that (f_i) is a bounded sequence in $L^p(\mathbb{R}^n)$. By Theorem 5.23 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 $L^p(\mathbb{R}^n)$. Thus from boundedness we obtain weak convergence for a subsequence.

(2) If $f_{i_j} \to f$ weakly in $L^p(\mathbb{R}^n)$, by Mazur's lemma there is a sequence of convex combinations that converges in $L^p(\mathbb{R}^n)$. Thus from weak convergence we obtain strong convergence for a subsequence of convex combinations.

(3) Strong convergence in $L^p(\mathbb{R}^n)$ implies almost everywhere convergence for a subsequence by Corollary 1.34. Thus from strong convergence we obtain almost everywhere convergence for a subsequence.

The following result is useful in identifying a weak limit in $L^p(\mathbb{R}^n)$.

Theorem 5.33. Let $1 \leqslant p < \infty$ and assume that $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$. If $f_i \to g$ almost everywhere in $\mathbb{R}^n$ as $i \to \infty$, then $f = g$ almost everywhere in $\mathbb{R}^n$.

Proof. By Lemma 5.32 there exists a sequence $(\widetilde{f}_i)$ of convex combinations

$$\widetilde{f}_i = \sum_{j=i}^{m_i} a_{i,j} f_j, \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=i}^{m_i} a_{i,j} = 1,$$

such that $\widetilde{f}_i \to f$ in $L^p(\mathbb{R}^n)$ as $i \to \infty$. By Corollary 1.34, there exists a subsequence, denoted again by $(\widetilde{f}_i)$, such that $\widetilde{f}_i \to f$ almost everywhere in $\mathbb{R}^n$. Since $f_i \to g$ almost everywhere in $\mathbb{R}^n$, we have

$$\widetilde{f}_i = \sum_{j=i}^{m_i} a_{i,j} f_j \to f$$

as $i \to \infty$. We conclude that $f = g$ almost everywhere in $\mathbb{R}^n$. □

Remark 5.34. if $f_i(x) \to f(x)$ as $i \to \infty$, then it does not follow for general convex combinations that

$$\sum_{j=1}^{m_i} a_{i,j} f_j(x) \xrightarrow{i \to \infty} f(x), \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=1}^{m_i} a_{i,j} = 1.$$

However, for tail convex combinations, we have

$$\sum_{j=i}^{m_i} a_{i,j} f_j(x) \xrightarrow{i\to\infty} f(x), \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=i}^{m_i} a_{i,j} = 1.$$

This is the advantage of the tail version of Mazur's lemma, see Lemma 5.32.

Corollary 5.35. Let $1 < p < \infty$ and assume that (f_i) is a bounded sequence in $L^p(\mathbb{R}^n)$. If $f_i \to f$ almost everywhere in $\mathbb{R}^n$, then $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$.

Proof. By Theorem 5.23 there exists a subsequence, still denoted by (f_i), and a function $g \in L^p(\mathbb{R}^n)$ such that $f_i \to g$ weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$. By Lemma 5.32 there exists a sequence $(\widetilde{f}_i)$ of convex combinations

$$\widetilde{f}_i = \sum_{j=i}^{m_i} a_{i,j} f_j, \quad \text{with} \quad a_{i,j} \geqslant 0 \quad \text{and} \quad \sum_{j=i}^{m_i} a_{i,j} = 1,$$

such that $\widetilde{f}_i \to g$ in $L^p(\mathbb{R}^n)$ as $i \to \infty$. Since $f_i \to f$ almost everywhere in $\mathbb{R}^n$ as $i \to \infty$, we have

$$\widetilde{f}_i = \sum_{j=i}^{m_i} a_{i,j} f_j \xrightarrow{i\to\infty} f.$$

We conclude that $f = g$ almost everywhere in $\mathbb{R}^n$ and $f_i \to f$ weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$. □

Remark 5.36. Let $1 < p < \infty$. Since $L^p(\mathbb{R}^n)$ is a uniformly convex Banach space, the Banach–Saks theorem which asserts that a weakly convergent sequence has a subsequence whose arithmetic means converge in the norm. Assume that a sequence $(f_i)_{i\in\mathbb{N}}$ converges to f weakly in $L^p(\mathbb{R}^n)$ as $i \to \infty$. Then there exists a subsequence $(f_{i_j})_{j\in\mathbb{N}}$ for which the arithmetic mean $\frac{1}{k}\sum_{j=1}^{k} f_{i_j}$ converges to f in $L^p(\mathbb{R}^n)$ as $k \to \infty$. The advantage of the Banach–Saks theorem compared to Mazur's lemma is that we can work with the arithmetic means instead of more general convex combinations

THE END

Bibliography

[1] V.I. Bogachev, Measure Theory 1 and 2, Springer 2007.

[2] A.M. Bruckner, J.B. Bruckner, and B.S. Thomson, Real Analysis, Prentice-Hall, 1997.

[3] D.L. Cohn, Measure theory (2nd edition), Birkhäuser, 2013.

[4] L.C. Evans and R.F. Gariepy, Measure Theory and Fine Properties of Functions, CRC Press, 1992.

[5] G.B. Folland, Real Analysis. Modern Techniques and Their Applications (2nd edition), John Wiley & Sons, 1999.

[6] I. Fonseca and G. Leoni, Modern Methods in the Calculus of Variations: L^p spaces, Springer 2010.

[7] F. Jones, Lebesgue Integration on Euclidean Space (revised edition), Jones and Bartlett Publishers, 2001.

[8] K.L. Kuttler, Modern Real Analysis, CRC Press, 1998.

[9] P. Mattila, Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability, Cambridge University Press, 1995.

[10] W. Rudin, Real and Complex Analysis, McGraw-Hill, 1986.

[11] D.A. Salamon, Measure and Integration, EMS, 2016,

[12] E. Stein and R. Sakarchi, Real Analysis: Measure Theory, Integration, and Hilbert Spaces, Princeton University Press, 2005.

[13] M.E. Taylor, Measure theory and integration, American Mathematical Society, 2006.

[14] R.L. Wheeden and A. Zygmund, Measure and Integral: An introduction to Real Analysis, Marcel Dekker, 1977.

[15] W.P. Ziemer, Modern Real Analysis, PWS Publishing Company, 1995.

[16] J. Yeh, Real Analysis, Theory of Measure and Integration (2nd edition), World Scientific, 2006.

Index