Real Analysis I: Notes
Henri Martikainen
University of Helsinki, Fall 2019

1. Contents

These notes supplement the notes by I. Holopainen [1]. We give an outline of the fall 2019 lectures here. Sometimes we simply refer to [1]. If we do something differently as in [1], we record the results and the proofs here.

1.1. *Remark.* We introduce here the following highly convenient notation. We denote $A \lesssim B$ if $A \le CB$ for some unimportant constant C. This means that C cannot depend on anything relevant like some important parameter ϵ. That is, C can e.g. be some uniform constant, or some constant depending on some fixed integrability exponent p. We can write $A \lesssim_\epsilon B$ to mean that $A \le C(\epsilon)B$ for some constant $C(\epsilon)$ that is now allowed to depend on some given parameter ϵ. We will also write $A \sim B$ if $A \lesssim B \lesssim A$.

In what follows $L^p = L^p(\mathbb{R}^n) = L^p(\mathbb{R}^n, \mathrm{d}x)$ – that is, we use the Lebesgue measure and **we operate in the whole of $\mathbb{R}^n$, unless we explicitly write $L^p(A)$ for some $A \subset \mathbb{R}^n$ or $L^p(X)$ for some measure space (X, μ).** Many results e.g. in $L^p(A)$ can be obtained by just setting $f = 0$ outside A and using the $\mathbb{R}^n$ results – but not everything can be obtained like this. We also write $\|f\|_p = \|f\|_{L^p(\mathbb{R}^n)}$, but explicitly write $\|f\|_{L^p(A)}$.

Moreover, $|A|$ denotes the Lebesgue measure of $A \subset \mathbb{R}^n$ and 1_A denotes the indicator function of the set A (often denoted by χ_A). Given $p \in [1, \infty]$ the dual exponent p' is defined via $1/p + 1/p' = 1$.

First week of lectures. Following [1]: Definition of L^p spaces, $1 \le p \le \infty$, obvious extension of the definition for $p \in (0, 1)$, Young's inequality, Hölder's inequality, triangle inequality of L^p spaces for $1 \le p \le \infty$ (Minkowski's inequality), completeness of L^p spaces. An alternative proof of Young's inequality below in Lemma 1.2.

1.2. **Lemma** (Young's inequality)**.** *Let $1 < p < \infty$. Then we have*

$$ab \le \frac{a^p}{p} + \frac{b^{p'}}{p'}, \qquad a, b \ge 0.$$

Proof. Define

$$h(x) = \frac{x^p}{p} + \frac{1}{p'} - x, \qquad x \ge 0,$$

and notice that, by elementary analysis (differentiation),

$$h(x) \ge h(1) = 0, \text{ i.e., } x \le \frac{x^p}{p} + \frac{1}{p'}.$$

Apply this with $x = ab^{1/(1-p)}$ to get

$$ab^{1/(1-p)} \le \frac{a^p}{p} b^{-p'} + \frac{1}{p'}.$$

Here we used that $p' = p/(p-1)$. Multiply both sides of this inequality with

$$b^{1-\frac{1}{1-p}} = b^{\frac{-p}{1-p}} = b^{p'}$$

to establish the desired result. $\square$

Second week of lectures. Sections 2-4 below.

Third week of lectures. Sections 5-7 below.

Fourth week of lectures. Section 8 below + Egorov's theorem (Thm 2.11) and Lusin's theorem (Thm 2.14) from [1] + absolute continuity of measures from [1] (Section 2.1).

Fifth week of lectures. A general version of the basic covering theorem (Thm 3.3 in [1]) (this is a more general version of Theorem 6.1 below), Vitali covering theorem (Thm 3.9 in [1]), properties of monotonic functions in $\mathbb{R}$ (Section 3.39 of [1]), definition of functions of bounded variation (beginning of Section 3.54).

Sixth week of lectures. Pages 51–59 of [1]: functions of bounded variation and all the main properties of absolutely continuous functions.

Seventh week of lectures. Theorem 3.79 and Theorem 3.83 from [1] (the final results regarding absolutely continuous functions). Section 9 below. Revision, time permitting.

2. APPROXIMATION BY CONTINUOUS FUNCTIONS AND CONTINUITY OF TRANSLATIONS

Let $C_c = C_c(\mathbb{R}^n)$ denote continuous and compactly supported functions (notice that Holopainen uses the subscript 0 instead of c).

2.1. **Lemma.** *Let $1 \le p < \infty$. Then C_c is dense in L^p. In other words, given $f \in L^p$ for every $\epsilon > 0$ there exists $g \in C_c$ so that $\|f - g\|_p < \epsilon$.*

Proof. We may assume that $f \ge 0$ (write $f = f_+ - f_-$) and that f is compactly supported (by DCT (dominated convergence theorem) $1_{B(0,M)} f \to f$ in L^p). In the course 'Measure and Integration' it is proved that there exists simple functions s_i so that $0 \le s_1 \le s_2 \le \ldots \le f$ and $f(x) = \lim_{i\to\infty} s_i(x)$. By DCT (or MCT) we have $\|s_i - f\|_p \to 0$. Thus, we may assume that f is itself a compactly supported simple function of the form $f = \sum_{i=1}^m a_i 1_{A_i}$. After this, we can clearly assume $f = 1_A$ for some bounded measurable set A (if we can approximate this by continuous functions, we can also approximate finite linear combinations).

Fix $\epsilon > 0$. Choose a compact K and and open G so that $K \subset A \subset G$ and $|G \setminus K| < \epsilon$ (see 'Measure and Integration' – the existence of these sets follows from the construction of the Lebesgue measure). We can now define the continuous function g approximating $f = 1_A$ explicitly:

$$g(x) = \frac{d(x, G^c)}{d(x, G^c) + D(x, K)}.$$

Remember that a mapping of the form $x \mapsto d(x, B)$ is 1-Lipschitz ($|d(x,B) - d(y,B)| \le |x-y|$) and so, in particular, continuous for all sets B. Notice also that the denominator in the definition of the function g is always strictly positive. Indeed, if $d(x, K) = 0$, then $x \in K$ (as K is closed) and so $x \in G$ and $d(x, G^c) > 0$ (as G open). Thus, g is continuous. Notice that $0 \le g \le 1$, $g(x) = 1$ if $x \in K$ and $g(x) = 0$ if $x \notin G$. Thus, $g - 1_A$ satisfies $|g - 1_A| \le 1$ and that $g(x) - 1_A(x) = 0$ unless $x \in G \setminus K$. We now simply get

$$\|g - f\|_p \le \Big(\int_{G \setminus K} 1 \, \mathrm{d}x \Big)^{1/p} = |G \setminus K|^{1/p} < \epsilon^{1/p},$$

and we are done. □

We give an elementary proof of the following very useful result, Theorem 2.29 in [1]. The proof in [1] uses Lusin's theorem and a lemma based on absolute continuity. However, heavier tools like that are not really required. Indeed, the result is almost trivial for C_c functions, and we can approximate general L^p functions by continuous functions using the above lemma.

2.2. **Lemma.** *Suppose $f \in L^p(\mathbb{R})$ for $1 \le p < \infty$. Then*

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

Proof. Fix $f \in L^p$ and denote the translation operator $\tau_y f(x) = f(x+y)$. Let $\epsilon > 0$ and choose $g \in C_c$ so that $\|f - g\|_p < \epsilon$. By translation invariance also $\|\tau_y f - \tau_y g\|_p = \|f - g\|_p < \epsilon$. Hence, by the triangle inequality for the L^p norm (Minkowski's inequality), it is enough to show that

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

Take an arbitrary sequence $y_k \to 0$. We may suppose that $|y_k| < 1$. Choose $M > 1$ so that $\operatorname{spt} g \subset B(0, M)$. Notice that if $g(x + y_k) \ne 0$, then $x + y_k \in B(0, M)$ and so $x \in B(0, 2M)$. Thus, we have

$$\|g - \tau_{y_k} g\|_p^p = \int_{B(0,2M)} |g(x) - g(x + y_k)|^p \, \mathrm{d}x.$$

As g is bounded, $|g(x) - g(x + y_k)| \le C$ and $C \in L^p(B(0, 2M))$), DCT gives

$$\lim_{k \to \infty} \int_{B(0,2M)} |g(x) - g(x+y_k)|^p \, \mathrm{d}x = \int_{B(0,2M)} \lim_{k \to \infty} |g(x) - g(x+y_k)|^p \, \mathrm{d}x = 0,$$

where we used the continuity of g with a fixed x. We are done. □

3. Convolution and L^p convergence of approximate identities

We define the convolution as in [1], but then we present somewhat more general approximation results involving the notion of 'approximation of identity'.

So, for $f, g \in L^1$ we define for $x \in \mathbb{R}^n$ the convolution

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

This is well-defined, since $\int_{\mathbb{R}^n} |f(y)g(x-y)|\,\mathrm{d}y < \infty$ for a.e. $x \in \mathbb{R}^n$. The latter follows from

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} |f(y)g(x-y)|\,\mathrm{d}y\,\mathrm{d}x = \int_{\mathbb{R}^n} |f(y)| \int_{\mathbb{R}^n} |g(x-y)|\,\mathrm{d}x\,\mathrm{d}y = \Big(\int_{\mathbb{R}^n} |f|\Big)\Big(\int_{\mathbb{R}^n} |g|\Big).$$

This used Fubini's theorem and translation invariance. Therefore, we have $f * g \in L^1$ and

$$\|f * g\|_1 \le \|f\|_1 \|g\|_1.$$

The following properties of the convolution are left as an exercise (here $f, g, h \in L^1$):

(1) $f * (g + h) = f * g + f * h$;
(2) $(\lambda f) * g = \lambda(f * g)$, $\lambda \in \mathbb{R}$;
(3) $f * g = g * f$;
(4) $f * (g * h) = (f * g) * h$;

3.1. *Remark.* It is important that the convolution of two functions can be defined in more generality. It is an exercise to prove the following result. Let $1 \le p, q, r \le \infty$ satisfy

$$\frac{1}{r} + 1 = \frac{1}{p} + \frac{1}{q}.$$

If $f \in L^p$ and $g \in L^q$ we have $f * g \in L^r$ and

$$\|f * g\|_r \le \|f\|_p \|g\|_q.$$

In particular, $f * g \in L^p$ if $f \in L^p$ and $g \in L^1$. Some of the calculations in the proof of Proposition 3.4 give a hint how to do this.

3.2. **Definition.** A family $\varphi_\epsilon \in L^1$, $\epsilon > 0$, is an approximate identity (as $\epsilon \to 0$) if the following conditions hold.

(1) We have $\int_{\mathbb{R}^n} \varphi_\epsilon = 1$ for all $\epsilon > 0$.
(2) We have $\sup_\epsilon \|\varphi_\epsilon\|_1 < \infty$.
(3) For every $\delta > 0$ we have

$$\lim_{\epsilon \to 0} \int_{|x| \ge \delta} |\varphi_\epsilon(x)|\,\mathrm{d}x = 0.$$

3.3. *Remark.* The following pointers regarding approximate identities are often helpful.

- Notice that if $\varphi_\epsilon \ge 0$, then (2) follows from (1). This is often the case.
- If $\operatorname{spt} \varphi_\epsilon \subset B(0, c(\epsilon))$, where $\lim_{\epsilon \to 0} c(\epsilon) = 0$, then (3) holds.

- If a fixed function $\eta \in L^1$ satisfies $\int \eta = 1$ and $\operatorname{spt} \eta \subset B(0,1)$, then $\eta_\epsilon := \frac{1}{\epsilon^n}\eta(x/\epsilon)$ is clearly an approximate identity. In fact, the condition $\operatorname{spt} \eta \subset B(0,1)$ is not needed (exercise).

Convolutions with approximate identities $f * \varphi_\epsilon$ are a very important way to approximate a given function $f \in L^p$ as $\epsilon \to 0$. Notice that by Remark 3.1 $f * \varphi_\epsilon$ is a well-defined L^p function if $f \in L^p$, $1 \le p < \infty$ (as $\varphi_\epsilon \in L^1$).

3.4. **Proposition.** *Let $1 \le p < \infty$, $f \in L^p$ and $(\varphi_\epsilon)_{\epsilon > 0}$ be an approximate identity. Then we have*

$$\|f - f * \varphi_\epsilon\|_p \to 0, \qquad \epsilon \to 0.$$

Proof. Fix $f \in L^p$. Using $\int_{\mathbb{R}^n} \varphi_\epsilon = 1$ and $f * \varphi_\epsilon = \varphi_\epsilon * f$ we write the pointwise identity

$$\begin{aligned} f(x) - f * \varphi_\epsilon(x) &= f(x) \int_{\mathbb{R}^n} \varphi_\epsilon(y)\,\mathrm{d}y - \int_{\mathbb{R}^n} f(x-y)\varphi_\epsilon(y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} [f(x) - f(x-y)]\varphi_\epsilon(y)\,\mathrm{d}y. \end{aligned}$$

For the moment let $p > 1$. We get using Hölder's inequality that

$$\begin{aligned} |f(x) - f * \varphi_\epsilon(x)| &\le \int_{\mathbb{R}^n} |f(x) - f(x-y)||\varphi_\epsilon(y)|^{1/p}|\varphi_\epsilon(y)|^{1/p'}\,\mathrm{d}y \\ &\le \Big(\int_{\mathbb{R}^n} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)|\,\mathrm{d}y\Big)^{1/p}\Big(\int_{\mathbb{R}^n} |\varphi_\epsilon(y)|\,\mathrm{d}y\Big)^{1/p'} \\ &\lesssim \Big(\int_{\mathbb{R}^n} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)|\,\mathrm{d}y\Big)^{1/p}, \end{aligned}$$

where the last step used that $\sup_\epsilon \|\varphi_\epsilon\|_1 \lesssim 1$. Therefore, we have

$$|f(x) - f * \varphi_\epsilon(x)|^p \lesssim \int_{\mathbb{R}^n} |f(x) - f(x-y)|^p|\varphi_\epsilon(y)|\,\mathrm{d}y,$$

which also clearly holds with $p = 1$. We integrate this over $x \in \mathbb{R}^n$, and use Fubini's theorem, to get that

$$\|f - f * \varphi_\epsilon\|_p^p \lesssim \int_{\mathbb{R}^n} |\varphi_\epsilon(y)| \int_{\mathbb{R}^n} |f(x) - f(x-y)|^p\,\mathrm{d}x\,\mathrm{d}y.$$

Let $\gamma > 0$. Using Lemma 2.2 we find $\delta > 0$ so that

$$\int_{\mathbb{R}^n} |f(x) - f(x-y)|^p\,\mathrm{d}x < \gamma$$

whenever $|y| < \delta$. Using property (3) of Definition 3.2 we find ϵ_0 so that

$$\int_{|y| \ge \delta} |\varphi_\epsilon(y)|\,\mathrm{d}y < \gamma$$

for all $\epsilon \le \epsilon_0$. For all $\epsilon \le \epsilon_0$ we therefore have

$$\int_{\mathbb{R}^n} |\varphi_\epsilon(y)| \int_{\mathbb{R}^n} |f(x) - f(x-y)|^p\,\mathrm{d}x\,\mathrm{d}y$$

$$
\begin{aligned}
&= \int_{|y|<\delta} |\varphi_\epsilon(y)| \int_{\mathbb{R}^n} |f(x) - f(x-y)|^p \, \mathrm{d}x \, \mathrm{d}y \\
&+ \int_{|y|\geq\delta} |\varphi_\epsilon(y)| \int_{\mathbb{R}^n} |f(x) - f(x-y)|^p \, \mathrm{d}x \, \mathrm{d}y \\
&\lesssim \gamma \int_{\mathbb{R}^n} |\varphi_\epsilon(y)| \, \mathrm{d}y + \|f\|_p^p \int_{|y|\geq\delta} |\varphi_\epsilon(y)| \, \mathrm{d}y.
\end{aligned}
$$

Recalling $\sup_\epsilon \|\varphi_\epsilon\|_1 \lesssim 1$ and $\int_{|y|\geq\delta} |\varphi_\epsilon(y)| \, \mathrm{d}y < \gamma$ we get that for all $\epsilon \leq \epsilon_0$ we have

$$\|f - f * \varphi_\epsilon\|_p^p \lesssim \gamma(1 + \|f\|_p^p).$$

□

4. Approximation by smooth functions in $L^p(U)$

$L^p(U)$ and $L^p_{\mathrm{loc}}(U)$. We will sometimes now work in

$$L^p(U) = \Big\{ f\colon U \to \mathbb{R}\colon \|f\|_{L^p(U)} := \Big(\int_U |f|^p \Big)^{1/p} < \infty \Big\}$$

for a given, fixed open set $U \subset \mathbb{R}^n$. The set U can be the whole space $\mathbb{R}^n$ – in particular, U need not be bounded. Define still the local L^p space

$$L^p_{\mathrm{loc}}(U) = \{ f\colon\ f \in L^p(V) \text{ for all open } V \subset\subset U \},$$

where $V \subset\subset U$ means that $\overline{V}$ is compact and $\overline{V} \subset U$. Notice that $L^p_{\mathrm{loc}}(U) \subset L^1_{\mathrm{loc}}(U)$ for all $1 \leq p \leq \infty$ as in each $V \subset\subset U$ we have by Hölder's inequality that

$$\int_V |f| \leq |V|^{1/p'} \Big(\int_V |f|^p \Big)^{1/p} < \infty.$$

That is why it is often convenient to state results for L^1_{loc} functions.

Differentiation and $C^k(U)$ spaces. Let $\alpha = (\alpha_1, \ldots, \alpha_n)$, where $\alpha_i \in \{0, 1, 2 \ldots\}$. This is called a multi-index and we set $|\alpha| = \sum_{i=1}^n \alpha_i$. Multi-indices are used mainly for notation related to differentiation: If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ then $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha^n}$, and

$$D^\alpha f(x) := \frac{\partial^{\alpha_1} \cdots \partial^{\alpha_n}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f(x) = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f(x),$$

where

$$\frac{\partial f(x)}{\partial x_j} = \lim_{h \to 0} \frac{f(x + he_j) - f(x)}{h}.$$

Here $(e_1, \ldots, e_n)$ is the standard basis of $\mathbb{R}^n$.

Define

$$
\begin{aligned}
C(U) = C^0(U) &= \{ f\colon U \to \mathbb{R}\colon\ f \text{ continuous} \}, \\
C^k(U) &= \{ f\colon U \to \mathbb{R}\colon\ D^\alpha f \in C(U) \text{ for } |\alpha| \leq k \}
\end{aligned}
$$

and

$$C^\infty(U) = \bigcap_{k=0}^\infty C^k(U).$$

Define also
$$C_c^k(U) = \{f \in C^k(U) \colon \operatorname{spt} f \subset U \text{ compact}\},$$
and define $C_c^\infty(U)$ analogously. Here the support $\operatorname{spt} f$ is taken in the whole of $\mathbb{R}^n$ so that a priori $\operatorname{spt} f \subset \overline{U}$, but we demand here that $\operatorname{spt} f \subset U$. In this case we say that 'f is compactly supported in U'.

Standard mollifier. In the exercise set 2 it is asked to show that $h \colon \mathbb{R} \to \mathbb{R}$ that is defined by $h(t) = e^{-1/t}$ for $t > 0$ and $h(t) = 0$ otherwise is a smooth function in $\mathbb{R}$ (i.e. indefinitely differentiable, $h \in C^\infty(\mathbb{R})$). We define $\eta \colon \mathbb{R}^n \to \mathbb{R}$ by setting
$$\eta(x) = Ch(1 - |x|^2)$$
for a constant $C > 0$ to be selected. Notice that $\eta \geq 0$, $\eta(x) = 0$ if $|x| \geq 1$, and that η is smooth – $\eta \in C_c^\infty(\mathbb{R}^n)$ (denoted by C_0^∞ by Holopainen). We let
$$C = \frac{1}{\int_{B(0,1)} h(1 - |x|^2)\,\mathrm{d}x}$$
so that, in addition, $\int \eta = \int_{B(0,1)} \eta = 1$. By Remark 3.3 we have that the family

$$\eta_\epsilon(x) := \frac{1}{\epsilon^n}\eta(x/\epsilon), \qquad \epsilon > 0, \tag{4.1}$$

is an approximate identity with $\eta_\epsilon(x) = 0$ if $|x| \geq \epsilon$. This particular family η_ϵ is the so-called 'standard mollifier', and is practical in many approximation arguments. In what follows η_ϵ denotes always this particular standard mollifier.

Smoothing of functions. If $f \in L^1_{\text{loc}} = L^1_{\text{loc}}(\mathbb{R}^n)$ (in particular if $f \in L^p_{\text{loc}}$) we can define
$$f^\epsilon(x) := f * \eta_\epsilon(x) = \int f(y)\eta_\epsilon(x - y)\,\mathrm{d}y$$
for **all** $x \in \mathbb{R}^n$. This is because
$$\int |f(y)\eta_\epsilon(x - y)|\,\mathrm{d}y \lesssim_\epsilon \int_{B(x,\epsilon)} |f| < \infty.$$
Thus, here we do not need arguments based on Fubini or anything else fancier, and we have a working definition for all x.

4.2. *Remark.* If $f \in L^1_{\text{loc}}(U)$ we can define $f^\epsilon(x)$ for only **all** $x \in U_\epsilon$, where
$$U_\epsilon = \{x \in U \colon\ d(x, \partial U) > \epsilon\}.$$
This is because $B(x, \epsilon) \subset\subset U$ for $x \in U_\epsilon$, so that $f \in L^1(B(x, \epsilon))$. Things are always a bit more technically demanding if we are working in the local spaces $L^1_{\text{loc}}(U)$, as then arguments can only be ran in compact sets that are **inside** U. That is, we always need to worry about such containments.

In the simpler situation when $f \in L^p(U)$ for some $1 \leq p \leq \infty$, we can again define $f^\epsilon(x)$ for all $x \in \mathbb{R}^n$. In this case you can interpret $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ by setting $f = 0$ outside U.

There is a rough philosophy: the convolution of two functions is as regular as the more regular function of the two.

4.3. **Theorem.** *Let $\epsilon > 0$. If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, we have $f^\epsilon \in C^\infty(\mathbb{R}^n)$ and*

$$D^\alpha f^\epsilon = f * D^\alpha \eta_\epsilon$$

for all multi-indices α.

Proof. Fix $x \in \mathbb{R}^n$ and $i \in \{1, \ldots, n\}$, and assume $|h| < \epsilon$. We can write

$$\frac{f^\epsilon(x + he_i) - f^\epsilon(x)}{h} = \frac{1}{\epsilon^n} \int_{B(x,2\epsilon)} \frac{1}{h} \left[\eta\Big(\frac{x - y + he_i}{\epsilon}\Big) - \eta\Big(\frac{x-y}{\epsilon}\Big) \right] f(y)\,\mathrm{d}y. \tag{4.4}$$

Notice that for all y we have

$$\begin{aligned} \frac{1}{h} \left[\eta\Big(\frac{x - y + he_i}{\epsilon}\Big) - \eta\Big(\frac{x-y}{\epsilon}\Big) \right] &= \frac{1}{\epsilon}\frac{1}{h/\epsilon} \left[\eta\Big(\frac{x - y}{\epsilon} + \frac{h}{\epsilon}e_i\Big) - \eta\Big(\frac{x-y}{\epsilon}\Big) \right] \\ &\to \frac{1}{\epsilon}\frac{\partial \eta}{\partial x_i}\Big(\frac{x-y}{\epsilon}\Big) = \epsilon^n \frac{\partial \eta_\epsilon}{\partial x_i}(x-y) \end{aligned}$$

as $h \to 0$. Hence, if we could pass the limit $h \to 0$ inside the integral in (4.4), we would get that

$$\frac{\partial f^\epsilon}{\partial x_i}(x) = \lim_{h \to 0} \frac{f^\epsilon(x + he_i) - f^\epsilon(x)}{h} = \int f(y) \frac{\partial \eta_\epsilon}{\partial x_i}(x-y)\,\mathrm{d}y = f * \frac{\partial \eta_\epsilon}{\partial x_i}(x).$$

The passage of the limit is justified by DCT as

$$\begin{aligned} \eta\Big(\frac{x - y}{\epsilon} + \frac{h}{\epsilon}e_i\Big) - \eta\Big(\frac{x-y}{\epsilon}\Big) &= \int_0^{h/\epsilon} \frac{d}{ds}\eta\Big(\frac{x-y}{\epsilon} + se_i\Big)\,\mathrm{d}s \\ &= \int_0^{h/\epsilon} \nabla\eta\Big(\frac{x-y}{\epsilon} + se_i\Big) \cdot e_i \,\mathrm{d}s, \end{aligned}$$

the absolute value of which is dominated by

$$\frac{h}{\epsilon}\|\nabla \eta\|_{L^\infty}.$$

Thus, the absolute value of the integrand in (4.4) is dominated by

$$\frac{1}{\epsilon}\|\nabla \eta\|_{L^\infty}|f|,$$

which is independent of the sequence variable h and belongs to $L^1(B(x, 2\epsilon))$, justifying the use of DCT. It is clear that we can repeat this argument to get $D^\alpha f^\epsilon = f * D^\alpha \eta_\epsilon$ for all multi-indices α.

To show that these are all continuous functions is a similar, but easier, argument. For example, for a fixed $x \in \mathbb{R}^n$, $\epsilon > 0$ and for $|h| < \epsilon$, we have

$$|f^\epsilon(x+h) - f^\epsilon(x)| \le \frac{1}{\epsilon^n}\int_{B(x,2\epsilon)} |f(y)| \left| \eta\Big(\frac{x-y+h}{\epsilon}\Big) - \eta\Big(\frac{x-y}{\epsilon}\Big) \right| \mathrm{d}y.$$

We can simply dominate (as η is bounded) the integrand by $C|f| \in L^1(B(x, 2\epsilon))$, and so the claim follows by using DCT as $h \to 0$ and the continuity of η. □

4.5. *Remark.* By small modifications we could prove $f^\epsilon \in C^\infty(U_\epsilon)$ if $f \in L^1_{\mathrm{loc}}(U)$.

4.6. **Theorem.** *Let $f \in L^p = L^p(\mathbb{R}^n)$, $1 \le p < \infty$. Then we have that*

$$\|f^\epsilon\|_p \le \|f\|_p$$

and

$$\lim_{\epsilon \to 0} \|f - f^\epsilon\|_p = 0.$$

Proof. By Remark 3.1 we have

$$\|f^\epsilon\|_p = \|f * \eta_\epsilon\|_p \le \|\eta_\epsilon\|_1 \|f\|_p = \|f\|_p.$$

That

$$\lim_{\epsilon \to 0} \|f - f^\epsilon\|_p = 0$$

follows from the fact that η_ϵ is an approximation of identity and Proposition 3.4. □

4.7. **Corollary.** *$C_c^\infty(U)$ is dense in $L^p(U)$, $1 \le p < \infty$.*

Proof. Fix $f \in L^p(U)$. For $j = 1, 2, \ldots$ define

$$V_j = B(0, j) \cap \{x \in U \colon d(x, \partial U) > 1/j\}.$$

By DCT we have $\|f 1_{V_j} - f\|_{L^p(U)} \to 0$ as $j \to \infty$. Let $\delta > 0$, and using the above choose $V_j \subset\subset U$ so that $\|f 1_{V_j} - f\|_{L^p(U)} < \delta$. Viewing $f 1_{V_j}$ as a locally integrable function defined in the whole $\mathbb{R}^n$, we know that $(f 1_{V_j})^\epsilon \in C^\infty(\mathbb{R}^n)$ for all $\epsilon > 0$. Notice that

$$(f 1_{V_j})^\epsilon(x) = \int_{V_j \cap B(x,\epsilon)} f(y) \eta_\epsilon(x - y)\,\mathrm{d}y.$$

Thus, $(f 1_{V_j})^\epsilon(x) = 0$ unless $V_j \cap B(x, \epsilon) \ne \emptyset$. This means that

$$\mathrm{spt}(f 1_{V_j}) \subset \{x \colon d(x, V_j) \le \epsilon\},$$

which is a compact subset of U if $\epsilon < 1/j$. Thus, we have $(f 1_{V_j})^\epsilon \in C_c^\infty(U)$ for $\epsilon < 1/j$. Using the above theorem we choose $\epsilon < 1/j$ so that

$$\|f 1_{V_j} - (f 1_{V_j})^\epsilon\|_{L^p(U)} = \|f 1_{V_j} - (f 1_{V_j})^\epsilon\|_p < \delta.$$

We have found $(f 1_{V_j})^\epsilon \in C_c^\infty(U)$ so that $\|f - (f 1_{V_j})^\epsilon\|_{L^p(U)} < 2\delta$. □

5. Interpolation

Let (X, μ) be a σ-finite measure space ($X = \bigcup_{i=1}^\infty X_i$, $\mu(X_i) < \infty$). This particular topic is completely general, and e.g. the translation invariance properties of the Lebesgue measure are not important here (unlike in the above convolution arguments). For $0 < p < \infty$ and a measurable $f \colon X \to \mathbb{R}$ define

$$\begin{aligned}
\|f\|_{L^p(X)} &= \Big(\int_X |f|^p \,\mathrm{d}\mu \Big)^{1/p}, \\
\|f\|_{L^{p,\infty}(X)} &= \sup_{\lambda > 0} \lambda \mu(\{x \in X \colon |f(x)| > \lambda\})^{1/p}, \\
\|f\|_{L^\infty(X)} &= \inf\{C \ge 0 \colon |f(x)| \le C \text{ for } \mu\text{-a.e. } x \in X\}, \\
\|f\|_{L^{\infty,\infty}(X)} &= \|f\|_{L^\infty(X)}.
\end{aligned}$$

The so-called weak-$L^p(X)$ – denoted $L^{p,\infty}(X)$ – consists of those f for which $\|f\|_{L^{p,\infty}(X)} < \infty$. If $f \in L^p(X)$ then for all $\lambda > 0$ we have

$$\mu(\{|f| > \lambda\}) = \frac{1}{\lambda^p} \int_{\{|f|>\lambda\}} \lambda^p \le \frac{1}{\lambda^p} \int |f|^p$$

from which it follows that $\|f\|_{L^{p,\infty}(X)} \le \|f\|_{L^p(X)} < \infty$. That is, we have the natural inclusion $L^p(X) \subset L^{p,\infty}(X)$.

5.1. **Theorem** (Marcinkiewicz interpolation theorem)**.** *Let (X, μ) and (Y, ν) be σ-finite measure spaces and let $0 < p_0 < p_1 \le \infty$. Let T be a* ***sublinear*** *operator defined on the space $L^{p_0}(X) + L^{p_1}(X)$ and taking values in the space of measurable functions on Y. Assume that there exists two constants A_0 and A_1 such that*

$$\begin{aligned} \|Tf\|_{L^{p_0,\infty}(Y)} &\le A_0 \|f\|_{L^{p_0}(X)}, \qquad f \in L^{p_0}(X), \\ \|Tf\|_{L^{p_1,\infty}(Y)} &\le A_1 \|f\|_{L^{p_1}(X)}, \qquad f \in L^{p_1}(X). \end{aligned}$$

Let $p \in (p_0, p_1)$ and write

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \qquad \theta \in (0,1).$$

Then we have

$$\|Tf\|_{L^p(Y)} \le 2\Big(\frac{p}{p-p_0} + \frac{p}{p_1 - p}\Big)^{1/p} A_0^{1-\theta} A_1^{\theta} \|f\|_{L^p(X)}.$$

5.2. *Remark.* Sublinearity means that we have the pointwise estimates

$$|T(f+g)| \le |Tf| + |Tg| \qquad \text{and} \qquad |T(\lambda f)| = |\lambda||Tf|,\ \lambda \in \mathbb{R}.$$

Marcinkiewicz interpolation theorem is an easy but very useful interpolation theorem. The good points are:

(1) We can assume only $L^q \to L^{q,\infty}$ type estimates at the endpoints $q \in \{p_0, p_1\}$ but conclude strong $L^p \to L^p$ estimates for $p_0 < p < p_1$.
(2) T does not need to be linear – this is important in what follows (T will e.g. be a so-called maximal function).

This theorem has a rather simple proof using the important identity

$$\int_X |f|^p \,\mathrm{d}\mu = p \int_0^\infty \lambda^{p-1} \mu(\{x \in X \colon |f(x)| > \lambda\}) \,\mathrm{d}\lambda, \qquad 0 < p < \infty. \tag{5.3}$$

The proof of this identity is left as an exercise. The weak point of the Marcinkiewicz interpolation theorem is that we cannot interpolate estimates like $L^{p_0} \to L^{q_0}$ and $L^{p_1} \to L^{q_1}$, but rather need to have $p_0 = q_0$ and $p_1 = q_1$. Such interpolation results do exist (the Riesz-Thorin interpolation theorem), but we will not cover those here.

Proof of Theorem 5.1. Assume $p_1 < \infty$ – the case $p_1 = \infty$ is an exercise. Let $f \in L^p$, $p_0 < p < p_1$. Fix some parameter $\lambda > 0$ related to the level sets of the form $\{|g| > \lambda\}$ appearing in (5.3) and fix also another technical parameter $\delta > 0$ (which we will later fix in a natural way to recover the claimed quantitative estimate).

Define $f_0 = f 1_{\{|f|>\delta\lambda\}}$ and $f_1 = f - f_0$. It is almost obvious that $f_0 \in L^{p_0}(X)$ (as $p_0 - p < 0$) and $f_1 \in L^{p_1}(X)$ (as $p_1 - p > 0$) – in particular, Tf is defined by assumption and we have by sublinearity that

$$|Tf| \le |Tf_0| + |Tf_1|.$$

Therefore, we have

$$\{|Tf| > \lambda\} \subset \{|Tf_0| > \lambda/2\} \cup \{|Tf_1| > \lambda/2\},$$

and so

$$\begin{aligned}
\nu(\{|Tf| > \lambda\}) &\le \nu(\{|Tf_0| > \lambda/2\}) + \nu(\{|Tf_1| > \lambda/2\}) \\
&\le \Big(\frac{\lambda}{2}\Big)^{-p_0} \|Tf_0\|_{L^{p_0,\infty}(Y)}^{p_0} + \Big(\frac{\lambda}{2}\Big)^{-p_1} \|Tf_1\|_{L^{p_1,\infty}(Y)}^{p_1} \\
&\le \Big(\frac{\lambda}{2}\Big)^{-p_0} A_0^{p_0} \|f_0\|_{L^{p_0}(X)}^{p_0} + \Big(\frac{\lambda}{2}\Big)^{-p_1} A_1^{p_1} \|f_1\|_{L^{p_1}(X)}^{p_1} \\
&= \Big(\frac{\lambda}{2}\Big)^{-p_0} A_0^{p_0} \int_{|f|>\delta\lambda} |f(x)|^{p_0}\,\mathrm{d}\mu(x) + \Big(\frac{\lambda}{2}\Big)^{-p_1} A_1^{p_1} \int_{|f|\le\delta\lambda} |f(x)|^{p_1}\,\mathrm{d}\mu(x).
\end{aligned}$$

In the last estimate we used the main assumption concerning the weak type estimates $L^{p_0}(X) \to L^{p_0,\infty}(Y)$ and $L^{p_1}(X) \to L^{p_1,\infty}(Y)$.

Using (5.3) we get that

$$\begin{aligned}
\|Tf\|_{L^p(Y)}^p &= p \int_0^\infty \lambda^{p-1} \nu(\{|Tf| > \lambda\})\,\mathrm{d}\lambda \\
&\le p(2A_0)^{p_0} \int_0^\infty \lambda^{p-1}\lambda^{-p_0} \int_{|f|>\delta\lambda} |f(x)|^{p_0}\,\mathrm{d}\mu(x)\,\mathrm{d}\lambda \\
&+ p(2A_1)^{p_1} \int_0^\infty \lambda^{p-1}\lambda^{-p_1} \int_{|f|\le\delta\lambda} |f(x)|^{p_1}\,\mathrm{d}\mu(x)\,\mathrm{d}\lambda = I + II.
\end{aligned}$$

If we want this generality of general measures, we now need Fubini's theorem with a general σ-finite measure. The proof is really different than in the Lebesgue case (see 'Measure and Integration'), and is given in 'Real Analysis II'. However, later in the lecture notes we will only need the interpolation in the Lebesgue case so you can also assume that μ is the Lebesgue measure. In any case, by some version of Fubini's theorem we have

$$\begin{aligned}
I &= p(2A_0)^{p_0} \int_X |f(x)|^{p_0} \int_0^{|f(x)|/\delta} \lambda^{p-p_0-1}\,\mathrm{d}\lambda\,\mathrm{d}\mu(x) \\
&= \frac{p(2A_0)^{p_0}}{p-p_0} \frac{1}{\delta^{p-p_0}} \int_X |f(x)|^p\,\mathrm{d}\mu(x)
\end{aligned}$$

and similarly

$$II = \frac{p(2A_1)^{p_1}}{p_1-p} \frac{1}{\delta^{p-p_1}} \int_X |f(x)|^p\,\mathrm{d}\mu(x).$$

Therefore, we have already proved that

$$\|Tf\|_{L^p(Y)}^p \le p\Big(\frac{(2A_0)^{p_0}}{p-p_0}\frac{1}{\delta^{p-p_0}} + \frac{(2A_1)^{p_1}}{p_1-p}\frac{1}{\delta^{p-p_1}}\Big)\|f\|_{L^p(X)}^p.$$

If we want to recover the exact claimed quantitative dependence on the various constants (which will not be important to us in what follows), it is now natural to fix δ so that

$$(2A_0)^{p_0}\frac{1}{\delta^{p-p_0}} = (2A_1)^{p_1}\frac{1}{\delta^{p-p_1}},$$

which gives

$$\delta = \frac{1}{2}A_0^{\frac{p_0}{p_1-p_0}}A_1^{-\frac{p_1}{p_1-p_0}}.$$

We then get

$$\|Tf\|_{L^p(Y)} \le (2A_0)^{p_0/p}\frac{1}{\delta^{1-p_0/p}}\Big(\frac{p}{p-p_0}+\frac{p}{p_1-p}\Big)^{1/p}\|f\|_{L^p(X)},$$

where

$$(2A_0)^{p_0/p}\frac{1}{\delta^{1-p_0/p}} = 2A_0^{1-\frac{p_0p_1-pp_1}{p_0p-pp_1}}A_1^{\frac{p_0p_1-pp_1}{p_0p-pp_1}}.$$

We are done after solving for θ, which gives the desired formula

$$\theta = \frac{1/p-1/p_0}{1/p_1-1/p_0} = \frac{p_0p_1-pp_1}{p_0p-pp_1}.$$

□

6. Maximal function estimates

For a locally integrable $f \in L^1_{\mathrm{loc}} = L^1_{\mathrm{loc}}(\mathbb{R}^n;\,\mathrm{d}x)$ define the (centred) Hardy–Littlewood maximal function

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

In practical arguments it is often convenient to use the following larger maximal function as well

$$x \mapsto \sup_{B\text{ open ball}}\frac{1_B(x)}{|B|}\int_B|f(y)|\,\mathrm{d}y.$$

Notice that if $x \in B = B(z,r)$, then $B \subset B(x,2r)$, and so (as $|B| \sim r^n \sim |B(x,2r)|$) we have

$$\sup_{B\text{ open ball}}\frac{1_B(x)}{|B|}\int_B|f(y)|\,\mathrm{d}y \lesssim Mf(x).$$

That is, these are pointwise comparable functions, and results that hold for one of them, also hold for the other. We can call this other one the 'non-centred maximal function' and denote it e.g. by $M_{nc}f(x)$.

The maximal function is of fundamental use in analysis as it has good mapping properties and it e.g. dominates many other operators pointwise. We will now prove the mapping properties.

6.1. **Theorem** (Basic covering theorem)**.** *Let $\mathcal{B}$ be a finite family of open (or closed) balls in $\mathbb{R}^n$. Then there exists pairwise disjoint balls $B_1, B_2, \ldots, B_m \in \mathcal{B}$ such that*

$$\bigcup_{B\in\mathcal{B}}B \subset \bigcup_{i=1}^m 3B_i.$$

Proof. Let $\mathcal{B} = \{U_j\}_{j=1}^N$, where $U_j = B(x_j, r_j)$. As this is a finite collection, by reordering we may assume that $r_1 \ge r_2 \ge \ldots \ge r_N$. Let $B_1 = U_1$, and then let B_2 be the biggest ball U_j so that $U_j \not\subset 3B_1$ (if it exists). Let then B_3 be the biggest ball U_j so that $U_j \not\subset 3B_1 \cup 3B_2$ (if it exists). We continue this selection process as long as possible – the process finishes after a finite, say m, number of steps. It follows from the construction directly that

$$\bigcup_{B \in \mathcal{B}} B \subset \bigcup_{i=1}^m 3B_i.$$

Importantly, the balls B_i, $i = 1, \ldots, m$, are disjoint. To see this, suppose that $B_{i_1} \cap B_{i_2} \ne \emptyset$ for some $1 \le i_1 < i_2 \le m$. As the radius of B_{i_1} is also larger than or equal to the radius of B_{i_2}, we must have (by triangle inequality) that $B_{i_2} \subset 3B_{i_1}$. But this is a contradiction with the selection process.

□

6.2. *Remark.* If $f \in L^1$ is non-trivial ($f \ne 0$ on a set of positive measure), then $Mf \notin L^1$. Indeed, in this case in some ball $B_R = B(0, R)$ we must have

$$\int_{B_R} |f| \gtrsim 1.$$

If $|x| > R$, then $B_R \subset B(x, 2|x|)$, and so

$$Mf(x) \ge \frac{1}{|B(x, 2|x|)|} \int_{B(x,2|x|)} |f| \gtrsim \frac{1}{|x|^n}.$$

Notice that

$$\int_{\mathbb{R}^n \setminus B(0,R)} |x|^{-n} \, \mathrm{d}x = \sum_{k=0}^\infty \int_{2^k R \le |x| < 2^{k+1} R} |x|^{-n} \, \mathrm{d}x \gtrsim \sum_{k=0}^\infty 1 = \infty.$$

Despite the previous remark, we do have the following result. It is typical in analysis that an operator does not map L^1 to L^1 but does map L^1 to $L^{1,\infty}$.

6.3. **Theorem.** *We have that $M \colon L^1(\mathbb{R}^n) \to L^{1,\infty}(\mathbb{R}^n)$ boundedly – i.e.,*

$$\|Mf\|_{L^{1,\infty}} \lesssim \|f\|_1.$$

Proof. Fix $f \in L^1$ and $\lambda > 0$. Define

$$\Omega_\lambda := \{x \in \mathbb{R}^n \colon Mf(x) > \lambda\}.$$

Let $K \subset \Omega_\lambda$ be an arbitrary compact set, and for every $x \in K$ choose (using the fact that $Mf(x) > \lambda$) a radius $r_x > 0$ and the related ball $U_x = B(x, r_x)$ so that

$$\frac{1}{|U_x|} \int_{U_x} |f| > \lambda.$$

As $\{U_x \colon x \in K\}$ is an open cover of K, we can use compactness to choose a finite subfamily $U_{x_1}, \ldots, U_{x_N}$ so that

$$K \subset \bigcup_{j=1}^N U_{x_j}.$$

By the basic covering theorem choose disjoint $B_1, \ldots, B_m \in \{U_{x_j} : j = 1, \ldots, N\}$ so that

$$K \subset \bigcup_{j=1}^{N} U_{x_j} \subset \bigcup_{i=1}^{m} 3B_i.$$

We now get

$$|K| \le \sum_{i=1}^{m} |3B_i| \lesssim \sum_{i=1}^{m} |B_i| \le \frac{1}{\lambda} \sum_{i=1}^{m} \int_{B_i} |f| \le \frac{1}{\lambda} \int_{\mathbb{R}^n} |f|.$$

As $K \subset \Omega_\lambda$ was an arbitrary compact subset, the same inequality holds with $|K|$ replaced by $|\Omega_\lambda|$, and we are done. □

6.4. **Corollary.** *For all $1 < p < \infty$ and $f \in L^p$ we have*

$$\|Mf\|_p \lesssim \|f\|_p.$$

Proof. As we have $\|Mf\|_{L^{1,\infty}} \lesssim \|f\|_1$ and the trivial estimate $\|Mf\|_\infty \le \|f\|_\infty$, the claim follows from Marcinkiewicz interpolation theorem. □

7. Lebesgue's differentiation theorem

7.1. **Theorem** (Lebesgue's differentiation theorem)**.** *For $f \in L^1_{\text{loc}}$ we have*

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

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

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

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

Proof. The latter claim follow from the first as

$$\frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) \, \mathrm{d}y - f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} [f(y) - f(x)] \, \mathrm{d}y,$$

and so it is enough to prove the first claim.

This is a local claim, so we can assume without loss of generality that $f \in L^1$ (enough to prove that the claim holds for every k and for a.e. $x \in B(0,k)$ – with a fixed k we can replace f by $f1_{B(0,2k)} \in L^1$).

There is a standard protocol to show almost everywhere convergence for integrable functions. It involves the following two steps: 1) show convergence in some appropriate dense subset; 2) prove the boundedness of the relevant maximal operator (depending on the problem at hand). In this case, the relevant maximal function is Mf, and we already know Theorem 6.3 – this gives us 2). But 1) is also clear, as the claim is obvious for continuous functions (which are dense). We now show how the standard protocol pieces these two facts together.

Let

$$\sigma_r f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y) - f(x)| \, \mathrm{d}y.$$

It is enough to show that

$$|\{x\colon \limsup_{r\to 0} \sigma_r f(x) > 0\}| = 0.$$

We fix an arbitrary $\lambda > 0$ and show that

$$|\{x\colon \limsup_{r\to 0} \sigma_r f(x) > \lambda\}| = 0,$$

which is enough. Let $\epsilon > 0$. Choose $g \in C_c$ so that

$$\|f - g\|_1 < \epsilon.$$

We know that because g is continuous we have

$$\lim_{r\to 0} \sigma_r g(x) = 0$$

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

$$\sigma_r f(x) \le \sigma_r(f-g)(x) + \sigma_r g(x)$$

we see that

$$\limsup_{r\to 0} \sigma_r f(x) \le \sup_{r>0} \sigma_r(f-g)(x) \le M(f-g)(x) + |f(x) - g(x)|.$$

Therefore, we have by Theorem 6.3 that

$$\begin{aligned}
&\Big|\Big\{x\colon \limsup_{r\to 0} \sigma_r f(x) > \lambda\Big\}\Big| \\
&\le \Big|\Big\{x\colon M(f-g)(x) > \frac{\lambda}{2}\Big\}\Big| + \Big|\Big\{x\colon |f(x)-g(x)| > \frac{\lambda}{2}\Big\}\Big| \\
&\le \frac{2}{\lambda}\big(\|M(f-g)\|_{L^{1,\infty}} + \|f-g\|_{L^{1,\infty}}\big) \lesssim \frac{1}{\lambda}\|f-g\|_1 < \frac{\epsilon}{\lambda}.
\end{aligned}$$

This ends the proof. □

7.2. *Remark.* Notice that Lebesgue's differentiation theorem implies that $|f(x)| \le Mf(x)$ for almost every x.

We present two immediate but important corollaries.

7.3. **Corollary.** *Let $f \in L^1([a,b])$ and define*

$$F(x) = \int_a^x f(y)\,\mathrm{d}y, \qquad x \in [a,b].$$

For almost every $x \in [a,b]$ we have

$$F'(x) = f(x).$$

Proof. Suppose $h > 0$. We have

$$\Big|\frac{F(x+h)-F(x)}{h} - f(x)\Big| \le \frac{1}{h}\int_x^{x+h} |f(y)-f(x)|\,\mathrm{d}y \le \frac{2}{2h}\int_{x-h}^{x+h} |f(y)-f(x)|\,\mathrm{d}y,$$

which goes, for almost every x, to 0 as $h \to 0+$ by Lebesgue's differentiation theorem. We can control the limit $\lim_{h\to 0-}$ similarly, and then the claim follows. □

7.4. **Corollary.** *Let $E \subset \mathbb{R}^n$ be measurable. Then for a.e. $x \in E$ we have*

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

and for a.e. $x \in E^c$ we have

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

Proof. Lebesgue's differentiation theorem applied to $1_E \in L^1_{\text{loc}}$ gives that

$$\lim_{r \to 0} \frac{|E \cap B(x,r)|}{|B(x,r)|} = \lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} 1_E(y)\,\mathrm{d}y = 1_E(x)$$

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

8. Pointwise convergence of approximate identities

We have already shown that if (φ_ϵ) is an approximate identity, then $f * \varphi_\epsilon \to f$ in L^p, $1 \le p < \infty$. We now study some related pointwise results and the case $p = \infty$.

If η_ϵ is the standard mollifier from (4.1), then we have

$$|f * \eta_\epsilon(x) - f(x)| \lesssim \frac{1}{\epsilon^n} \int_{B(x,\epsilon)} |f(y) - f(x)|\,\mathrm{d}y \sim \frac{1}{|B(x,\epsilon)|} \int_{B(x,\epsilon)} |f(y) - f(x)|\,\mathrm{d}y.$$

Thus, it follows from Lebesgue's differentiation theorem that if $f \in L^1_{\text{loc}}$, then $f * \eta_\epsilon(x) \to f(x)$ for a.e. $x \in \mathbb{R}^n$. We can prove this more generally. If $\varphi \in L^1$ with $\int \varphi = 1$, then (see exercises) we know that the scaled functions $\varphi_\epsilon(x) = \frac{1}{\epsilon^n}\varphi(x/\epsilon)$ form an approximate identity. Under certain additional assumptions on φ we will show that $f * \varphi_\epsilon(x) \to f(x)$ for a.e. $x \in \mathbb{R}^n$.

We say that φ is radial if its values $\varphi(x)$ only depend on $|x|$. This means that $\varphi(x) = r(|x|)$ for some $r\colon [0,\infty] \to \mathbb{R}$. We say that a radial φ is decreasing if $|x| \ge |y|$ implies $\varphi(x) \le \varphi(y)$. Our aim is to show that $f * \varphi_\epsilon(x) \to f(x)$ for a.e. $x \in \mathbb{R}^n$ if $f \in L^p$, $1 \le p \le \infty$, and φ is radial, non-negative and decreasing. Notice that the function η in the standard mollifier satisfies these assumptions, but that the function η is even compactly supported and smooth, which are not required in the general theorem.

The scheme for showing this is the same as in the Lebesgue's differentiation theorem (the case $f \in L^\infty$ being a bit special): 1) show convergence in some appropriate dense subset; 2) prove the boundedness of the relevant maximal operator (depending on the problem at hand). We will start with 2).

8.1. **Proposition.** *Suppose $f \in L^p$, $1 \le p \le \infty$. Let $\varphi \in L^1$ be radial, non-negative and decreasing and $\varphi_\epsilon(x) = \frac{1}{\epsilon^n}\varphi(x/\epsilon)$ be the related approximate identity. Then for all $\epsilon > 0$ we have*

$$|f * \varphi_\epsilon(x)| \le \|\varphi\|_1 Mf(x)$$

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

Proof. Due to the properties of φ, we can approximate φ by the special 'simple' functions of the form $\sum_i a_i 1_{B(0,r_i)}$, where $a_i, r_i > 0$. Indeed, if $\varphi(x) = r(|x|)$ for a positive and decreasing function r defined on $[0, \infty]$, we can approximate pointwise using for each $j = 1, 2, \ldots$ the functions

$$\varphi_j(x) = \sum_{i=1}^{\infty} [r(i2^{-j}) - r((i+1)2^{-j})] 1_{B(0,i2^{-j})}(x).$$

By the monotone convergence theorem it is enough to prove that

$$|f * (\varphi_j)_\epsilon(x)| \le \|\varphi_j\|_1 Mf(x).$$

So, for notational convenience, we can assume that $\varphi(x) = \sum_i a_i 1_{B(0,r_i)}(x)$, $a_i, r_i > 0$. Then we have

$$f * \varphi_\epsilon(x) = \frac{1}{\epsilon^n} \int_{\mathbb{R}^n} f(x-y)\varphi(y/\epsilon)\,\mathrm{d}y = \int_{\mathbb{R}^n} f(x-\epsilon y)\varphi(y)\,\mathrm{d}y,$$

and so

$$\begin{aligned}
|f * \varphi_\epsilon(x)| &\le \sum_i a_i \int_{B(0,r_i)} |f(x-\epsilon y)|\,\mathrm{d}y \\
&= \sum_i a_i \frac{1}{\epsilon^n} \int_{B(x,\epsilon r_i)} |f(y)|\,\mathrm{d}y \\
&= \sum_i a_i |B(0,r_i)| \frac{1}{|B(x,\epsilon r_i)|} \int_{B(x,\epsilon r_i)} |f(y)|\,\mathrm{d}y \\
&\le \Big(\sum_i a_i |B(0,r_i)|\Big) Mf(x) = \|\varphi\|_1 Mf(x).
\end{aligned}$$

□

For 1) we need the following lemma that is of independent interest.

8.2. **Lemma.** *Let $(\varphi_\epsilon)_\epsilon$ be a general approximate identity and $g \in L^\infty$ be continuous at the point x. Then we have*

$$g(x) = \lim_{\epsilon \to 0} g * \varphi_\epsilon(x).$$

Proof. Suppose $\gamma > 0$. Choose $\delta > 0$ so that $|g(z) - g(x)| < \gamma$ whenever $|z - x| < \delta$. Then using the properties of approximate identities choose ϵ_0 so that

$$\int_{|y| \ge \delta} |\varphi_\epsilon(y)|\,\mathrm{d}y < \gamma$$

for all $\epsilon \le \epsilon_0$. Then we have

$$\begin{aligned}
|g(x) - g * \varphi_\epsilon(x)| &\le \int_{\mathbb{R}^n} |g(x-y) - g(x)||\varphi_\epsilon(y)|\,\mathrm{d}y \\
&= \Big(\int_{|y|<\delta} + \int_{|y| \ge \delta}\Big) |g(x-y) - g(x)||\varphi_\epsilon(y)|\,\mathrm{d}y = I + II.
\end{aligned}$$

We have for all ϵ that

$$I \le \gamma \int_{\mathbb{R}^n} |\varphi_\epsilon(y)|\,\mathrm{d}y \lesssim \gamma.$$

For all $\epsilon \le \epsilon_0$ we have

$$II \lesssim \|g\|_\infty \int_{|y|\ge\delta} |\varphi_\epsilon(y)|\,\mathrm{d}y \le \gamma\|g\|_\infty.$$

We are done. $\square$

8.3. *Remark.* If $g \in L^\infty$ is uniformly continuous (e.g. $g \in C_c$) then the above proof gives that

$$\|g - g * \varphi_\epsilon\|_\infty \to 0.$$

8.4. **Theorem.** *Suppose $f \in L^p$, $1 \le p \le \infty$. Let $\varphi \in L^1$ be radial, non-negative and decreasing and $\varphi_\epsilon(x) = \frac{1}{\epsilon^n}\varphi(x/\epsilon)$ be the related approximate identity. Then we have*

$$f(x) = \lim_{\epsilon\to 0} f * \varphi_\epsilon(x)$$

for almost every x.

Proof. Having proved the above results, the proof is now (in the case $1 \le p < \infty$) completely analogous to the proof of Lebesgue's differentiation theorem. Indeed, define

$$\sigma_\epsilon f(x) = |f * \varphi_\epsilon(x) - f(x)|$$

and the related maximal function

$$Af(x) = \sup_{\epsilon>0} |f * \varphi_\epsilon(x)|.$$

By Proposition 8.1 we have $Af(x) \lesssim Mf(x)$. We know by Theorem 6.3 and Corollary 6.4 that A maps boundedly as follows: $A\colon L^1 \to L^{1,\infty}$ and $A\colon L^p \to L^p$, $1 < p < \infty$. In particular, we have $A\colon L^p \to L^{p,\infty}$ for all $1 \le p < \infty$. For $1 \le p < \infty$ we are now in the position to run the exact same 'standard argument' as in Lebesgue's differentiation theorem. We give the details to make this absolutely clear.

So suppose $f \in L^p$, $1 \le p < \infty$. It is enough to show that for an arbitrary $\lambda > 0$ we have

$$|\{x\colon \limsup_{\epsilon\to 0} \sigma_\epsilon f(x) > \lambda\}| = 0.$$

Let $\gamma > 0$. Choose $g \in C_c$ so that

$$\|f - g\|_p < \gamma.$$

By Lemma 8.2 we have

$$\lim_{\epsilon\to 0} \sigma_\epsilon g(x) = 0$$

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

$$\sigma_\epsilon f(x) \le \sigma_\epsilon(f-g)(x) + \sigma_\epsilon g(x)$$

we see that

$$\limsup_{\epsilon\to 0} \sigma_\epsilon f(x) \le \sup_{\epsilon>0} \sigma_\epsilon(f-g)(x) \le A(f-g)(x) + |f(x) - g(x)|.$$

Therefore, as $\|Ah\|_{L^{p,\infty}} \lesssim \|h\|_p$ we have

$$\left|\left\{x\colon \limsup_{\epsilon\to 0} \sigma_\epsilon f(x) > \lambda\right\}\right|$$

$$\le \left|\left\{x\colon A(f-g)(x) > \frac{\lambda}{2}\right\}\right| + \left|\left\{x\colon |f(x)-g(x)| > \frac{\lambda}{2}\right\}\right|$$
$$\le \left(\frac{\lambda}{2}\right)^{-p}\left(\|A(f-g)\|_{L^{p,\infty}}^p + \|f-g\|_{L^{p,\infty}}^p\right) \lesssim \frac{1}{\lambda^p}\|f-g\|_p^p < \frac{\gamma^p}{\lambda^p}.$$

This ends the proof in the case $1 \le p < \infty$.

Let now $p = \infty$ and $f \in L^\infty$. Fix an arbitrary $r > 0$. We will show that

$$f(x) = \lim_{\epsilon\to 0} f * \varphi_\epsilon(x)$$

for almost every $x \in B(0,r)$ – which is enough. Define $f_1 = f1_{B(0,r+1)}$ and $f_2 = f - f_1$. As $f_1 \in L^1$ we know by the first part of the proof that

$$f(x) = f_1(x) = \lim_{\epsilon\to 0} f_1 * \varphi_\epsilon(x)$$

for almost every $x \in B(0,r)$. Thus, it is enough to show that

$$\lim_{\epsilon\to 0} f_2 * \varphi_\epsilon(x) = 0$$

for almost every $x \in B(0,r)$. Write

$$f_2 * \varphi_\epsilon(x) = \int_{\mathbb{R}^n} 1_{B(0,r+1)^c}(x-y) f(x-y)\varphi_\epsilon(y)\,\mathrm{d}y.$$

If $x \in B(0,r)$ and $y \in B(0,1)$, then $x - y \in B(0, r+1)$ and so $1_{B(0,r+1)^c}(x-y) = 0$. Thus, we get

$$|f_2 * \varphi_\epsilon(x)| \le \|f\|_\infty \int_{\mathbb{R}^n\setminus B(0,1)} |\varphi_\epsilon(y)|\,\mathrm{d}y.$$

As $(\varphi_\epsilon)_\epsilon$ is an approximate identity, we know that $\lim_{\epsilon\to 0}\int_{|y|\ge 1}|\varphi_\epsilon(y)|\,\mathrm{d}y = 0$. Thus, we are done. $\square$

Poisson kernel and Dirichlet problem. As an application, we define the Poisson kernel $P \in L^1$ by setting

$$P(x) = \frac{C}{(1+|x|^2)^{(n+1)/2}}, \qquad x \in \mathbb{R}^n.$$

Notice that this is indeed in L^1 as

$$\int_{B(0,1)} P(x)\,\mathrm{d}x + \sum_{k=0}^{\infty}\int_{2^k\le|x|<2^{k+1}} P(x)\,\mathrm{d}x \lesssim 1 + \sum_{k=0}^{\infty} 2^{-k} < \infty.$$

Then we select the constant C so that $\int P = 1$. Thus, as $P \in L^1$ is non-negative, radial and decreasing, we know that for $f \in L^p$ we have pointwise almost everywhere that

$$\lim_{t\to 0} u(x,t) = f(x),$$

where $u(x,t) := f * P_t(x)$ and $P_t(x) = \frac{1}{t^n}P(x/t)$. Thus, the function u defined in the upper half-space

$$\mathbb{R}^{n+1}_+ := \{(x,t)\colon\ x \in \mathbb{R}^n, t > 0\}$$

has the 'boundary values' $f(x)$. If $1 \le p < \infty$, we also know that

$$\int_{\mathbb{R}^n} |u(x,t) - f(x)|^p\,\mathrm{d}x \to 0$$

as $t \to 0$. It can be shown that u is harmonic in $\mathbb{R}^{n+1}_+$ – that is, we have that the Laplace vanishes:

$$\Delta u(x,t) = \Big(\frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2} + \frac{\partial^2}{\partial t^2}\Big) u(x,t) = 0, \qquad (x,t) \in \mathbb{R}^{n+1}_+.$$

Thus, we have that u solves the boundary value problem (Dirichlet problem) $\Delta u = 0$ in $\mathbb{R}^{n+1}_+$ and $u = f$ in $\partial \mathbb{R}^{n+1}_+ = \mathbb{R}^n$ (in the above sense $\lim_{t\to 0} u(x,t) = f(x)$ pointwise almost everywhere). We have presented a way to study the Dirichlet problem with 'rough' boundary data – indeed, the boundary data f is only an L^p function. In addition, note still that by Proposition 8.1 we have

$$\sup_{t>0} |u(x,t)| \le Mf(x).$$

9. Weak derivatives and Sobolev spaces

L^p spaces are some of the most fundamental function spaces in modern analysis. They are important on their own, but also serve as a basic building block for more complicated spaces, such as Sobolev spaces, which appear naturally in the theory of partial differential equations. We give a very brief introduction to these spaces now.

In what follows $U \subset \mathbb{R}^n$ is open.

9.1. **Definition.** Let $f \in L^1_{\text{loc}}(U)$ and α be a multi-index. If there exists $g \in L^1_{\text{loc}}(U)$ so that the integration by parts type formula

$$\int_U f D^\alpha \varphi = (-1)^{|\alpha|} \int_U g\varphi$$

holds for all $\varphi \in C_c^\infty(U)$, then we denote $g = D^\alpha f$ and call g the α^{th}-weak partial derivative of f.

In the exercises we show that if $f \in L^1_{\text{loc}}(U)$ is such that

$$\int f\varphi = 0$$

for all $\varphi \in C_c^\infty(U)$, then $f = 0$ almost everywhere. It follows that that the weak derivates are unique, if they exist.

9.2. **Example.** Let $U = (0,2) \subset \mathbb{R}$ and

$$f(x) = \begin{cases} x, & \text{if } 0 < x \le 1, \\ 2, & \text{if } 1 < x < 2. \end{cases}$$

If $\varphi \in C_c^\infty(U)$, then

$$\int_0^2 f\varphi' = \int_0^1 x\varphi' - 2\varphi(1) = -\int_0^1 \varphi - \varphi(1).$$

Aiming for a contradiction suppose that there exists $g = D^1 f \in L^1_{\text{loc}}(U)$. Choose now a sequence of functions (φ_j) so that $\varphi_j \in C_c^\infty(U)$, $\varphi_j(1) = 1$, $0 \le \varphi_j \le 1$ and $\varphi_j(x) \to 0$, when $x \in (0,2)$, $x \ne 1$. We now have by DCT that

$$1 = \lim_{j\to\infty} \varphi_j(1) = \lim_{j\to\infty} \left(\int_0^2 g\varphi_j - \int_0^1 \varphi_j \right) = 0,$$

which is a contradiction.

If, however, we define

$$f(x) = \begin{cases} x, & \text{if } 0 < x \le 1, \\ 1, & \text{if } 1 < x < 2, \end{cases}$$

then we have that the weak derivative $D^1 f$ exists and

$$D^1 f(x) = \begin{cases} 1, & \text{if } 0 < x \le 1, \\ 0, & \text{if } 1 < x < 2. \end{cases}$$

To see this, notice that for all $\varphi \in C_c^\infty(U)$ we have

$$\begin{aligned} \int_0^2 f(x)\varphi'(x)\,\mathrm{d}x &= \int_0^1 x\varphi'(x)\,\mathrm{d}x + \int_1^2 \varphi'(x)\,\mathrm{d}x \\ &= -\int_0^1 \varphi(x)\,\mathrm{d}x + \varphi(1) - \varphi(1) = -\int_0^1 \varphi(x)\,\mathrm{d}x. \end{aligned}$$

Thus, an angle is fine, but an actual jump is too much for the weak derivative to exist.

9.3. **Definition.** Let $1 \le p < \infty$ and $k = 1, 2, \ldots$. We say that $f \in W^{k,p}(U)$ if $f \in L^p(U)$ has weak-derivatives $D^\alpha f \in L^p(U)$ for every multi-index α with $|\alpha| \le k$. We norm this space with the norm

$$\|f\|_{W^{k,p}(U)} := \sum ||D^\alpha f||_{L^p(U)},$$

where we agree $D^0 f = f$. These are called Sobolev spaces.

9.4. **Theorem.** *$W^{k,p}(U)$ is a Banach space.*

Proof. For convenience, let $k = 1$. Let (f_j) be a Cauchy sequence in $W^{1,p}(U)$. Then (f_j) and $(\partial_i f_j)$, $1 \le i \le n$, are Cauchy sequences in $L^p(U)$, which we know to be a Banach space. Thus, there exists $g, g_1, \ldots, g_n \in L^p(U)$ so that $f_j \to g$ and $\partial_i f_j \to g_i$, $1 \le i \le n$, in $L^p(U)$. It is now enough to show that the weak derivative $\partial_i g$ exists and $\partial_i g = g_i$. Let $\varphi \in C_c^\infty(U)$ and $\operatorname{spt} \varphi \subset V \subset\subset U$. Then we have

$$\begin{aligned} \int_V g\partial_i\varphi &= \int_V (g - f_j)\partial_i\varphi + \int_V f_j\partial_i\varphi \\ &= \int_V (g - f_j)\partial_i\varphi - \int_V (\partial_i f_j)\varphi \\ &= \int_V (g - f_j)\partial_i\varphi - \int_V (\partial_i f_j - g_i)\varphi - \int_V g_i\varphi. \end{aligned}$$

The claim follows by letting $j \to \infty$, using Hölder's inequality and $|V| < \infty$. $\square$

We end our brief study of Sobolev spaces by illustrating that the fact that derivatives belong to L^p implies (in some situations) that $f \in L^q$ for $q > p$.

We assume $n \ge 2$ in what follows.

9.5. **Definition.** If $1 \le p < n$, the Sobolev conjugate of p is

$$p^* = \frac{np}{n-p}.$$

Notice that
$$\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}$$
and so $p^* > p$. Let us see how this exponent comes up naturally. Fix $1 \le p < n$. Suppose that there is $1 \le q < \infty$ and a constant C so that
$$\|\varphi\|_q \le C\|\nabla\varphi\|_p$$
for all $\varphi \in C_c^\infty(\mathbb{R}^n)$. Fix $\varphi \in C_c^\infty(\mathbb{R}^n)$ so that $\varphi \not\equiv 0$. Define for all $\lambda > 0$ the function φ_λ by setting $\varphi_\lambda(x) = \varphi(\lambda x)$. We have
$$\left(\int |\varphi(\lambda x)|^q \,\mathrm{d}x\right)^{1/q} \le C\left(\int |\lambda\nabla\varphi(\lambda x)|^p \,\mathrm{d}x\right)^{1/p} = C\lambda\left(\int |\nabla\varphi(\lambda x)|^p \,\mathrm{d}x\right)^{1/p}.$$
By change of variable this gives
$$\lambda^{-n/q}\|\varphi\|_q \le C\lambda^{1-n/p}\|\nabla\varphi\|_p$$
and so
$$\|\varphi\|_q \le C\lambda^{1-n/p+n/q}\|\nabla\varphi\|_p.$$
If $1 - n/p + n/q \ne 0$, then $\lambda^{1-n/p+n/q} \to 0$, when $\lambda \to 0$ or $\lambda \to \infty$, and we must have $\varphi \equiv 0$ – a contradiction. Thus, if the inequality
$$\|\varphi\|_q \le C\|\nabla\varphi\|_p$$
holds for all $\varphi \in C_c^\infty(\mathbb{R}^n)$, we must at least have $1 - n/p + n/q = 0$ and so
$$q = \frac{np}{n-p} = p^*.$$

We now prove that this inequality is, indeed, valid. Denote the closure of $C_c^\infty(U)$ in $W^{k,p}(U)$ by $W_0^{k,p}(U)$. In general, we have $W^{k,p}(U) \ne W_0^{k,p}(U)$. However, in the special $U = \mathbb{R}^n$ this is true (which we do not show here).

9.6. **Theorem** (The Sobolev inequality)**.** *Assume $1 \le p < n$. Then we have*
$$\|f\|_{p^*} \lesssim \|\nabla f\|_p$$
for all $f \in W^{1,p}(\mathbb{R}^n) = W_0^{1,p}(\mathbb{R}^n)$.

Proof. We begin with the case $p = 1$ – notice that $1^* = n/(n-1)$. Assume $\varphi \in C_c^\infty(\mathbb{R}^n)$. For all $1 \le i \le n$ and $x \in \mathbb{R}^n$ we have
$$\begin{aligned}|\varphi(x)| = |\varphi(x_1, \ldots, x_i, \ldots, x_n)| &= \left|\int_{-\infty}^{x_i} \partial_i\varphi(x_1, \ldots, t_i, \ldots, x_n)\,dt_i\right| \\ &\le \int_{-\infty}^{\infty} |\nabla\varphi(x_1, \ldots, t_i, \ldots, x_n)|\,dt_i.\end{aligned}$$

Thus, there holds
$$|\varphi(x)|^{\frac{n}{n-1}} \le \prod_{i=1}^n \left[\int_{-\infty}^{\infty} |\nabla\varphi(x_1, \ldots, t_i, \ldots, x_n)\,\mathrm{d}t_i\right]^{\frac{1}{n-1}}.$$

Integrating both sides with respect to $x_1 \in (-\infty, \infty)$ we get

$$\begin{aligned}\int_{-\infty}^{\infty} |\varphi|^{\frac{n}{n-1}} \, \mathrm{d}x_1 &\le \int_{-\infty}^{\infty} \prod_{i=1}^{n} \left[\int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_i \right]^{\frac{1}{n-1}} \mathrm{d}x_1 \\ &= \left[\int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_1 \right]^{\frac{1}{n-1}} \int_{-\infty}^{\infty} \prod_{i=2}^{n} \left[\int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_i \right]^{\frac{1}{n-1}} \mathrm{d}x_1.\end{aligned}$$

Using the generalised Hölder's inequality (from the exercises) with $\sum_{i=2}^{n} 1/1 = n-1 = \frac{1}{1/(n-1)}$ we get

$$\begin{aligned}\int_{-\infty}^{\infty} \prod_{i=2}^{n} \left[\int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_i \right]^{\frac{1}{n-1}} \mathrm{d}x_1 &= \left\| \prod_{i=2}^{n} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_i \right\|_{L^{\frac{1}{n-1}}_{x_1}}^{\frac{1}{n-1}} \\ &\le \left[\prod_{i=2}^{n} \left\| \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_i \right\|_{L^1_{x_1}} \right]^{\frac{1}{n-1}} \\ &= \left[\prod_{i=2}^{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}t_i \right]^{\frac{1}{n-1}}.\end{aligned}$$

Thus, we have

$$\int_{-\infty}^{\infty} |\varphi|^{\frac{n}{n-1}} \, dx_1 \le \left[\int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_1 \right]^{\frac{1}{n-1}} \prod_{i=2}^{n} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}t_i \right]^{\frac{1}{n-1}}.$$

We now integrate the obtained inequality over $x_2 \in (-\infty, \infty)$ to get

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi|^{\frac{n}{n-1}} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \le \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}t_2 \right]^{\frac{1}{n-1}} \int_{-\infty}^{\infty} \prod_{\substack{i=1 \\ i\ne 2}}^{n} I_i^{\frac{1}{n-1}} \, \mathrm{d}x_2,$$

where

$$I_1 = \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_1 \qquad \text{and} \qquad I_i = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}t_i, \quad i = 3, \ldots, n.$$

Applying the generalised Hölder's inequality as above we get

$$\begin{aligned}&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\varphi|^{\frac{n}{n-1}} \, \mathrm{d}x_1 \, \mathrm{d}x_2 \\ &\le \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}t_2 \right]^{\frac{1}{n-1}} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}t_1 \, \mathrm{d}x_2 \right]^{\frac{1}{n-1}} \\ &\times \prod_{i=3}^{n} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\nabla\varphi| \, \mathrm{d}x_1 \, \mathrm{d}x_2 \, \mathrm{d}t_i \right]^{\frac{1}{n-1}}.\end{aligned}$$

Continuing we eventually get

$$\int_{\mathbb{R}^n} |\varphi|^{\frac{n}{n-1}} \le \left[\int_{\mathbb{R}^n} |\nabla\varphi| \right]^{\frac{n}{n-1}}.$$

But this is the desired result for $\varphi \in C_c^\infty(\mathbb{R}^n)$ – that is, we have

$$\|\varphi\|_{L^{1^*}(\mathbb{R}^n)} \le \|\nabla\varphi\|_{L^1(\mathbb{R}^n)}.$$

Let then $f \in W^{1,1}(\mathbb{R}^n) = W_0^{1,1}(\mathbb{R}^n)$, and choose $\varphi_j \in C_c^\infty(\mathbb{R}^n)$ so that

$$\|f - \varphi_j\|_{W^{1,1}(\mathbb{R}^n)} \to 0.$$

By what we have proved we have

$$\|\varphi_i - \varphi_j\|_{L^{1^*}(\mathbb{R}^n)} \le \|\nabla\varphi_i - \nabla\varphi_j\|_{L^1(\mathbb{R}^n)}.$$

Thus, we have that (φ_j) is a Cauchy sequence in the space $L^{1^*}(\mathbb{R}^n)$, and so $f \in L^{1^*}(\mathbb{R}^n)$ and $\|f - \varphi_j\|_{L^{1^*}(\mathbb{R}^n)} \to 0$. Thus, we get

$$\|f\|_{L^{1^*}(\mathbb{R}^n)} = \lim_{j\to\infty} \|\varphi_j\|_{L^{1^*}(\mathbb{R}^n)} \le \lim_{j\to\infty} \|\nabla\varphi_j\|_{L^1(\mathbb{R}^n)} = \|\nabla f\|_{L^1(\mathbb{R}^n)}.$$

Let now $1 < p < n$. Assume again $\varphi \in C_c^\infty(\mathbb{R}^n)$. We apply the $p = 1$ case to the $W_0^{1,1}(\mathbb{R}^n)$ mapping $g = |\varphi|^\gamma$, where $\gamma > 1$ is to be selected. Now $\nabla g = \gamma|\varphi|^{\gamma-1}\nabla\varphi$ and so

$$\left(\int |\varphi|^{\frac{\gamma n}{n-1}}\right)^{\frac{n-1}{n}} \le \gamma \int |\varphi|^{\gamma-1}|\nabla\varphi| \le \gamma \left(\int |\varphi|^{\frac{(\gamma-1)p}{p-1}}\right)^{\frac{p-1}{p}} \|\nabla\varphi\|_{L^p(\mathbb{R}^n)}.$$

In the last step we used Hölder's inequality. Choosing

$$\gamma = \frac{p(n-1)}{n-p} > 1$$

we get

$$\frac{\gamma n}{n-1} = p^* = \frac{(\gamma-1)p}{p-1}.$$

Therefore, we have

$$\|\varphi\|_{L^{p^*}(\mathbb{R}^n)} = \left(\int_{\mathbb{R}^n} |\varphi|^{p^*}\right)^{1/p^*} = \left(\int_{\mathbb{R}^n} |\varphi|^{p^*}\right)^{\frac{n-1}{n} - \frac{p-1}{p}} \le \frac{p(n-1)}{n-p}\|\nabla\varphi\|_{L^p(\mathbb{R}^n)},$$

which is the desired result for smooth, compactly supported functions. The approximation argument is the same as in the $p = 1$ case. $\square$

We also now right away get the following version with an open set $U \subset \mathbb{R}^n$.

9.7. **Theorem.** *Assume $1 \le p < n$. Then we have*

$$\|f\|_{L^{p^*}(U)} \lesssim \|\nabla f\|_{L^p(U)}$$

for all $f \in W_0^{1,p}(U)$.

Proof. As $f \in W_0^{1,p}(U)$ there exists $\varphi_j \in C_c^\infty(U)$ so that $\|f - \varphi_j\|_{W^{1,p}(U)} \to 0$. We can set $\varphi_j = 0$ outside U so that $\varphi_j \in C_c^\infty(\mathbb{R}^n)$ and apply the previous result to φ_j. The claim follows with the same Cauchy sequence argument as above. $\square$

Thus, we have $W_0^{1,p}(U) \subset L^{p^*}(U)$ and $W^{1,p}(\mathbb{R}^n) \subset L^{p^*}(\mathbb{R}^n)$, even though only the inclusion to $L^p(U)$ (or $L^p(\mathbb{R}^n)$) was obvious from the definition. For $p = n$ and $p > n$ we have different type of inclusions. Appropriate inclusions also hold for $W_0^{k,p}(U)$ with a general k, of course.

References

[1] I. Holopainen, Real Analysis I, Lecture notes, University of Helsinki, 2017.

(H.M.) Department of Mathematics and Statistics, University of Helsinki, P.O.B. 68, FI-00014 University of Helsinki, Finland
E-mail address: henri.martikainen@helsinki.fi