Advanced Real Analysis

Edited and typeset by David McCormick
Based upon lectures by José Luis Rodrigo

University of Warwick

Autumn 2010

Preface

These notes are primarily based on the lectures for the course MA4J0 Advanced Real Analysis, given in the autumn of 2010 by Dr José Luis Rodrigo at the University of Warwick. I have embellished the material as lectured to include slightly more detail and exposition, and in one or two places I have reordered things a little to make the text flow a bit better.

In addition, I have included the material in José's handwritten notes which was not lectured. These sections are marked with a star both in the title and the contents page, and are not examinable (at least in 2010/2011). In particular, section 2.8 on the Lebesgue differentiation theorem, section 3.5 on the Fourier transform of tempered distributions, section 3.6 on Sobolev spaces, part of section 3.7 on fundamental solutions, and all of section 4 on the Hilbert transform were not lectured, and are not examinable.

I would very much appreciate being told of any errors or oddities in these notes, the responsibility for which is mine alone and no reflection on José's excellent lectures. Any corrections may be sent by email to `d.s.mccormick@warwick.ac.uk`; please be sure to include the version number and date given below.

David McCormick, University of Warwick, Coventry
Version 0.1 of January 17, 2011

Edition History

Version 0.1	January 17, 2011	Initial release for proofreading.

Contents

Introduction

Just over two hundred years ago, Joseph Fourier revolutionised the world of mathematics by writing down a solution to the heat equation by means of decomposing a function into a sum of sines and cosines. The need to understand these so-called Fourier series gave birth to analysis as we know it today: what's amazing is that the process of understanding Fourier series goes on, and Fourier analysis is still a fruitful area of research.

In this course we aim to give an introduction to the classical theory of Fourier analysis. There are four chapters, which cover Fourier series, the Fourier transform, distribution theory, and the Hilbert transform respectively. (Note that the starred sections are not examinable in 2010/2011.)

Some of the principal questions which serve as motivation for the study of Fourier analysis are as follows:

Fourier Series Let $f\colon [0, 2\pi] \to \mathbb{R}$ be a 2π-periodic function, and let its n^{th} Fourier coefficient be given by

$$\hat{f}(n) := \frac{1}{2\pi} \int_0^{2\pi} f(y) e^{-iny} \, \mathrm{d}y.$$

Can you recover f as $\sum_{n=-N}^{N} \hat{f}(n) e^{inx}$? If so, how does it converge, and under what conditions?

Fourier Transform Let $f\colon \mathbb{R} \to \mathbb{R}$ be any function, and define its Fourier transform $\hat{f}\colon \mathbb{R} \to \mathbb{C}$ by

$$\hat{f}(y) := \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{-iyx} \, \mathrm{d}x.$$

How, if possible, can we rebuild f from $\hat{f}$? Does $T_N f(x) := \int_{-\infty}^{+\infty} \hat{f}(s) \chi_{[-N,N]}(s) e^{ixs} \, \mathrm{d}s$ converge to $f(x)$? If so, how does it converge, and under what conditions? Whether $T_N f$ converges or not depends on the dimension; C. Fefferman won a Fields medal for that discovery.

Distributions In order to study Fourier series and Fourier transforms in full generality, we will need tools from the theory of distributions. For example, if we try and take the Fourier transform of $f(x) = e^{ix}$, we get that

$$\hat{f}(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ix} e^{-iyx} \, \mathrm{d}x,$$

so that $\hat{f}(y) = 0$ for $y \neq 0$, and $\hat{f}(0) = +\infty$, but in such a way that $\int_{\mathbb{R}^n} \hat{f}(y) \, \mathrm{d}y = 1$. Such an $\hat{f}$ isn't really a function: we thus need to generalise the notion of function to a distribution.

Hilbert Transform The central ideas of this course are all linked by the Hilbert transform:

$$Hf(x) = \text{p. v.} \int_{-\infty}^{+\infty} \frac{f(y)}{x - y} \, \mathrm{d}y$$

which uses techniques from complex analysis, Fourier series, singular integrals, and PDEs (the solution of the Laplacian). Let us now show – *heuristically!* – how $T_n f$ is related to the Hilbert transform Hf introduced above:

$$\begin{aligned}
T_N f(x) &= \int_{-\infty}^{+\infty} \hat{f}(s)\chi_{[-N,N]}(s)e^{ixs}\,\mathrm{d}s \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(y)e^{iys}\,\mathrm{d}y\right)\chi_{[-N,N]}(s)e^{ixs}\,\mathrm{d}s \\
&= \int_{-N}^{N} \left(\int_{-\infty}^{+\infty} f(y)e^{-iys}\quad rdy\right)e^{ixs}\,\mathrm{d}s \\
&= \int_{-\infty}^{+\infty} f(y)\left(\int_{-N}^{N} e^{i(x-y)s}\,\mathrm{d}s\right)\mathrm{d}y && \text{by Fubini} \\
&= \int_{-\infty}^{+\infty} f(y)\frac{1}{i(x-y)}\; e^{i(x-y)s}\big|_{-N}^{N}\;\mathrm{d}y && \text{if } x \neq y \\
&= \frac{1}{i}\int_{-\infty}^{+\infty} \frac{f(y)}{x-y}\left(e^{iN(x-y)} - e^{-iN(x-y)}\right)\mathrm{d}y \\
&= \frac{2i}{i}\int_{-\infty}^{+\infty} \frac{f(y)}{x-y}\sin N(x-y)\,\mathrm{d}y \\
&= 2H(\tilde{f})
\end{aligned}$$

for some modification $\tilde{f}$ of f. So as $N \to \infty$ for $T_N f$, the result is linked to $H\tilde{f}$; that is, the convergence of $T_N f$ and H is linked!

Books

The sections on Fourier series, the Fourier transform and the Hilbert transform are based on the book of Duoandikoetxea [Duo], while the section on distribution theory is based on the book of Friedlander and Joshi [Fri&Jos]. Both books are readable yet clear introductions to the subject.

For an introduction to Fourier series, Fourier transforms and their applications to differential equations, the books of Folland [Fol] and Stein and Shakarchi [Ste&Sha] are pitched below the level of the course and will be useful as background reading.

For further reading in Fourier analysis, the books of Grafakos [GraCl] and [GraMo] are comprehensive yet very readable, and are highly recommended; alternatively, the classic books by Stein [SteHA] and [SteSI] are excellent reference works. For background in the functional analysis and distribution theory involved, Rudin's book [Rud] is a nice introduction to the subject (which includes topological vector spaces), while Yosida's book [Yos] is a comprehensive reference, though it is perhaps a little outdated now.

1 Fourier Series

1.1 Basic Definitions and the Dirichlet Kernel

We will consider Fourier series on $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$; that is, a function $f\colon \mathbb{T} \to \mathbb{R}$ is a 2π-periodic function on $\mathbb{R}$ (or on $[-\pi, \pi]$ depending on your point of view). It will be convenient to abuse notation at various points and consider the domain of such functions to be $[0, 2\pi]$ or $[-\pi, \pi]$ or similar, as appropriate. We also define $C(\mathbb{T})$ to be the set of continuous functions on $[-\pi, \pi]$ which are 2π-periodic, $L^1(\mathbb{T})$ to be the set of L^1 functions on $[-\pi, \pi]$ which are 2π-periodic, and so on.

Definition 1.1 (Fourier coefficients)**.** *Given $f\colon \mathbb{T} \to \mathbb{R}$, define the n^{th}* Fourier coefficient *by*

$$\hat{f}(n) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y) e^{-iny} \, \mathrm{d}y.$$

Define the basis functions $e_n(x) := e^{inx}$. Then the e_n are orthogonal with respect to the L^2 inner product: for $f, g \in L^2(\mathbb{T})$ define their inner product by

$$\langle f, g \rangle := \int_{-\pi}^{\pi} f(x) \overline{g(x)} \, \mathrm{d}x.$$

We thus observe that

$$\hat{f}(n) = \frac{1}{2\pi} \langle f, e_n \rangle;$$

that is, $\hat{f}(n)$ is the proejction of f onto e_n.

From this, we note that $\hat{f}$ is naturally defined when $f \in L^1(\mathbb{T})$, but that $\hat{f}$ also makes sense when $f \in L^2(\mathbb{T})$. Furthermore, since $|e^{-iny}| \leq 1$, we have

$$|\hat{f}(n)| \leq \frac{1}{2\pi} \|f\|_{L^1}.$$

Theorem 1.2 (Riemann–Lebesgue lemma)**.** *If $f \in L^1(\mathbb{T})$, then $|\hat{f}(n)| \to 0$ as $n \to \pm\infty$.*

Proof. For $a \in \mathbb{R}$, we define $f_a(x) := f(x - a)$. Its Fourier coefficient is given by

$$\begin{aligned} \hat{f}_a(n) &= \frac{1}{2\pi} \int_0^{2\pi} f_a(y) e^{-iny} \, \mathrm{d}y \\ &= \frac{1}{2\pi} \int_0^{2\pi} f(y - a) e^{-iny} \, \mathrm{d}y \end{aligned}$$

We now let $z = y - a$; the limits of integration do not change as f is 2π-periodic, so we obtain:

$$\begin{aligned} \hat{f}_a(n) &= \frac{1}{2\pi} \int_0^{2\pi} f(z) e^{-in(z+a)} \, \mathrm{d}z \\ &= \frac{e^{-ina}}{2\pi} \frac{1}{2\pi} \int_0^{2\pi} f(z) e^{-inz} \, \mathrm{d}z \\ &= e^{-ina} \hat{f}(n). \end{aligned}$$

Now, choose $a = \pi/n$. Then $e^{-ina} = -1$; for such a, we have

$$\begin{aligned}
2|\hat{f}(n)| &= |\hat{f}(n) - e^{-ina}\hat{f}(n)| \\
&= |\hat{f}(n) - \hat{f}_a(n)| \\
&= \left| \frac{1}{2\pi} \int_0^{2\pi} f(y)e^{-iny}\,\mathrm{d}y - \frac{1}{2\pi}\int_0^{2\pi} f(y-a)e^{-iny}\,\mathrm{d}y \right| \\
&= \frac{1}{2\pi}\left| \int_0^{2\pi} \left(f(y) - f(y-a)\right) e^{-iny}\,\mathrm{d}y \right| \\
&\leq \int_0^{2\pi} |f(y) - f(y-a)|\,\mathrm{d}y.
\end{aligned}$$

Note that $a \to 0$ as $n \to +\infty$.

If $f \in C^0(\mathbb{T})$, then $|f(y) - f(y - \frac{\pi}{n})| \to 0$ for every $y \in [0, 2\pi]$. Hence by the Dominated Convergence Theorem,

$$\lim_{n\to+\infty} \int_0^{2\pi} |f(y) - f(y-a)|\,\mathrm{d}y = 0.$$

In general if $f \in L^1(\mathbb{T})$, there exists $g \in C^0(\mathbb{T}) \cap L^1(\mathbb{T})$ such that $\|f - g\|_{L^1} < \varepsilon/2$. Take K large enough so that $|\hat{g}(k)| < \varepsilon/2$ whenever $|k| \geq K$; then

$$|\hat{f}(k)| \leq |\hat{f}(k) - \hat{g}(k)| + |\hat{g}(k)| \leq \|f - g\|_{L^1} + \frac{\varepsilon}{2} < \varepsilon. \qquad \square$$

The fundamental question is whether we can recover f as a Fourier series. Define

$$S_N f(x) = \sum_{k=-N}^{N} \hat{f}(k)e^{ikx}.$$

Is it true that $S_N f(x) \to f(x)$ as $N \to \infty$? In general, that is far too much to hope for. For certain kinds of f convergence is assured; but we will see that some functions f are so weird that $S_N f(x)$ diverges for every $x \in \mathbb{T}$!

In order to study the convergence of Fourier series, it will be helpful to rewrite $S_N f$ as a particular kind of integral known as a *convolution*. Notice that

$$\begin{aligned}
S_N f(x) &= \sum_{k=-N}^{N} \hat{f}(k)e^{ikx} \\
&= \sum_{k=-N}^{N} \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)e^{-iky}\,\mathrm{d}y\, e^{ikx} \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) \sum_{k=-N}^{N} e^{ik(x-y)}\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) D_N(x-y)\,\mathrm{d}y \\
&= \frac{1}{2\pi} f * D_N,
\end{aligned}$$

where $D_N(t) = \sum_{k=-N}^{N} e^{ikt}$ is the Dirichlet kernel. Here $f * g$ means the convolution of f and g, which is defined as

$$f * g(x) = \int_0^{2\pi} f(y)g(x-y)\,\mathrm{d}y \underbrace{=}_{\substack{\text{change of}\\ \text{variables}}} \int_0^{2\pi} f(x-y)g(y)\,\mathrm{d}y.$$

The Dirichlet kernel is a somewhat awkward sum; the following result shows that we can reduce it to a single quotient of sines:

Lemma 1.3. $D_N(x) := \displaystyle\sum_{k=-N}^{N} e^{ikx} = \frac{\sin(N+\frac{1}{2})x}{\sin\frac{x}{2}}.$

Proof. We compute:

$$\begin{aligned}
D_N(x) &= \sum_{k=-N}^{N} e^{ikx} \\
&= e^{-iNx} \sum_{k=0}^{2N} e^{ikx} \\
&= e^{-iNx} \frac{1-e^{i(2N+1)x}}{1-e^{ix}} && \text{using the geometric series formula} \\
&= \frac{e^{-iNx} - e^{i(N+1)x}}{1-e^{ix}} \\
&= \frac{e^{-i(N+\frac{1}{2})x} - e^{i(N+\frac{1}{2})x}}{e^{-ix/2} - e^{ix/2}} && \text{multiplying top and bottom by } e^{-ix/2} \\
&= \frac{\frac{e^{-i(N+\frac{1}{2})x} - e^{i(N+\frac{1}{2})x}}{2i}}{\frac{e^{-ix/2} - e^{ix/2}}{2i}} \\
&= \frac{\sin(N+\frac{1}{2})x}{\sin\frac{x}{2}}.
\end{aligned}$$

□

Before we move on, let us observe that

$$\int_{-\pi}^{\pi} D_N(y)\,\mathrm{d}y = \int_{-\pi}^{\pi} \sum_{k=-N}^{N} e^{iky}\,\mathrm{d}y = 2\pi.$$

1.2 Convergence and Divergence

Using the Dirichlet kernel, we can go on to prove results about when the Fourier series $S_N f$ converges to f. The first three results rely *only* on the Riemann–Lebesgue lemma and do not require any more complicated results.

Theorem 1.4 (Convergence of $S_N f$ is local)**.** *Let $f \in L^1(\mathbb{T})$. Suppose that f is 0 in a neighbourhood of x; that is, there exists $\delta > 0$ such that $f(y) = 0$ for all $y \in (x-\delta, x+\delta)$. Then $S_N f(x) \to 0$.*

This result is significant in that while $\hat{f}(n)$ depends on the values of f *globally* — that is, to calculate $\hat{f}(n)$ we need to know f everywhere — the convergence of $S_N f(x)$ only depends on a local neighbourhood of x.

Proof. We compute:

$$\begin{aligned}
S_N f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) D_N(x-y)\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y) D_N(y)\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]} f(x-y) D_N(y)\,\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{[-\pi,\pi]\setminus[-\delta,\delta]} \frac{f(x-y)}{\sin\frac{y}{2}} \sin(N+\tfrac{1}{2})y\,\mathrm{d}y.
\end{aligned}$$

Now, $y \mapsto \frac{f(x-y)}{\sin\frac{y}{2}}$ is an L^1 function on $S := [-\pi,\pi]\setminus[-\delta,\delta]$, so

$$\begin{aligned}
S_N f(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \chi_S(y)\frac{f(x-y)}{\sin\frac{y}{2}}\frac{e^{i(N+\frac{1}{2})y}-e^{-i(N+\frac{1}{2})y}}{2i}\,\mathrm{d}y. \\
&= \frac{1}{2\pi}\frac{1}{2i}\int_{-\pi}^{\pi}\chi_S(y)\frac{f(x-y)}{\sin\frac{y}{2}}e^{iy/2}e^{iNy}\,\mathrm{d}y - \frac{1}{2\pi}\frac{1}{2i}\int_{-\pi}^{\pi}\chi_S(y)\frac{f(x-y)}{\sin\frac{y}{2}}e^{-iy/2}e^{-iNy}\,\mathrm{d}y.
\end{aligned}$$

Setting $g(y) := \chi_S(y)\frac{f(x-y)}{\sin\frac{y}{2}}e^{iy/2}$, and $h(y) := \chi_S(y)\frac{f(x-y)}{\sin\frac{y}{2}}e^{-iy/2}$, we see that the two terms are nothing but Fourier coefficients of g and h, so that

$$S_N f(x) = \frac{1}{2\pi}\hat{g}(-N) - \hat{h}(N) \to 0$$

by the Riemann-Lebesgue lemma. □

The next result represents pretty much the minimal hypotheses you need to ensure that $S_N f(x)$ converges to $f(x)$:

Theorem 1.5 (Dini's convergence theorem)**.** *Let $f \in L^1(\mathbb{T})$. Suppose that there exists $\delta > 0$ such that*

$$\int_{|t|<\delta}\left|\frac{f(x-t)-f(x)}{t}\right|\,\mathrm{d}t < +\infty.$$

Then $S_N f(x) \to f(x)$.

Recall that if $f \in C^1$, then

$$\lim_{t\to 0}\left|\frac{f(x-t)-f(x)}{t}\right|$$

exists and is bounded for all x; in which case, taking $\delta = \pi$,

$$\int_{|t|<\pi}\left|\frac{f(x-t)-f(x)}{t}\right|\,\mathrm{d}t \le 2\pi\|f'\|_{L^\infty} < +\infty.$$

Proof of theorem 1.5. We compute

$$
\begin{aligned}
|S_N f(x) - f(x)| &= \left| \frac{1}{2\pi} \int f(x-y) D_N(y)\,\mathrm{d}y - \frac{1}{2\pi} \int f(x) D_N(y)\,\mathrm{d}y \right| \\
&= \left| \frac{1}{2\pi} \int \frac{f(x-y) - f(x)}{|y|} \sin(N+\tfrac{1}{2})y \frac{|y|}{\sin(y/2)}\,\mathrm{d}y \right| \\
&= \frac{1}{2\pi} \left| \frac{1}{2i} \int_{-\pi}^{\pi} \frac{f(x-y)-f(x)}{|y|} \frac{|y|}{\sin(y/2)} \left(e^{iy/2} e^{iNy} - e^{-iy/2} e^{-iNy} \right)\,\mathrm{d}y \right| \\
&= \frac{1}{2\pi} \cdot \frac{1}{2} \Bigg| \int_{-\pi}^{\pi} \underbrace{\frac{f(x-y)-f(x)}{|y|} \frac{|y|}{\sin(y/2)} e^{iy/2}}_{=:g(y)} e^{iNy}\,\mathrm{d}y \\
&\qquad - \int_{-\pi}^{\pi} \underbrace{\frac{f(x-y)-f(x)}{|y|} \frac{|y|}{\sin(y/2)} e^{-iy/2}}_{=:h(y)} e^{-iNy}\,\mathrm{d}y \Bigg|.
\end{aligned}
$$

To complete the proof, it suffices to show that $g, h \in L^1(\mathbb{T})$. For g (noting that $|g| = |h|$, we need that

$$\int_{-\pi}^{\pi} \frac{|f(x-y) - f(x)|}{|y|} \frac{|y|}{|\sin(y/2)|} |e^{iy/2}|\,\mathrm{d}y < +\infty.$$

We split the integral into the regions where $|y| < \delta$, and $\delta < |y| < \pi$, as in the previous proof. For $\delta < |y| < \pi$, we have that $\frac{1}{|\sin(y/2)|} \leq M$, so

$$\int_{\delta<|y|<\pi} \frac{|f(x-y)-f(x)|}{|\sin(y/2)|} |e^{iy/2}|\,\mathrm{d}y \leq M \int_{\delta<|y|<\pi} |f(x-y) - f(x)|\,\mathrm{d}y \leq 2M \|f\|_{L^1}.$$

For $|y| < \delta$, we observe that $\frac{|y|}{|\sin(y/2)|} \leq 2$, so that

$$\int_{|y|<\delta} \frac{|f(x-y)-f(x)|}{|y|} \frac{|y|}{|\sin(y/2)|} |e^{iy/2}|\,\mathrm{d}y \leq \int_{|y|<\delta} \frac{|f(x-y)-f(x)|}{|y|} \cdot 2\,\mathrm{d}y < +\infty,$$

by assumption. □

Observe that for Dini's theorem to hold, it is in fact enough to have that there exist constants $R > 0$, $\alpha \in (0, 1]$ and $C > 0$ such that whenever $|y| \leq R$, we have

$$|f(x-y) - f(y)| \leq C|y|^\alpha.$$

Such functions are called *α-Hölder continuous*. (The definition of α-Hölder continuous is often stated without the restriction that $|x - y| \leq R$; the R is strictly only needed when the domain is non-compact.) Note that if $f \in C^1(\mathbb{T})$, then f is automatically 1-Hölder (also known as Lipschitz).

The next theorem tells us that if f is of bounded variation — loosely speaking, if f does not oscillate too much — then $S_N f(x)$ converges to the average of the left and right limits of f at x:

Theorem 1.6 (Jordan's criterion)**.** *Let $f \in L^1(\mathbb{T})$ be of bounded variation. Then*

$$S_N f(x) \to \frac{f(x^+) + f(x^-)}{2},$$

where $f(x^+) = \lim_{h\to 0^+} f(x+h)$, and $f(x^-) = \lim_{h\to 0^+} f(x-h)$.

Indeed, the theorem holds (and the same proof works) if f is only BV in a neighbourhood of x.

To prove Jordan's criterion, we first recall some facts about BV functions. Recall that $f\colon [a,b] \to \mathbb{R}$ is of bounded variation if

$$\sup\left\{\sum_{i=1}^{n} |f(t_i) - f(t_{i-1})| : a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b, n \in \mathbb{N}\right\} < +\infty.$$

Recall that the *total variation* of f is defined to be

$$T_f(x) := \sup\left\{\sum_{i=1}^{n} |f(x_j) - f(x_{j-1})| : a = x_0 < x_1 < \cdots < x_n = x\right\}$$

where the sup is taken over all partitions of $[a,b]$. Thus, f is of bounded variation on $[a,b]$ if and only if $T_f(x)$ is bounded on $[a,b]$. Furthermore, given any function $f \in \mathrm{BV}([a,b])$, we may write

$$f(x) = \frac{1}{2}(T_f(x) + f(x)) - \frac{1}{2}(T_f(x) - f(x));$$

observe that $f^+(x) := \frac{1}{2}(T_f(x) + f(x))$ and $f^-(x) := \frac{1}{2}(T_f(x) - f(x))$ are both monotone increasing functions. That is, a function is of bounded variation if and only if it is the difference of two monotone increasing functions.

We also recall the mean value formula for integrals:

Lemma 1.7 (Mean value formula for integrals)**.** *Let $\phi\colon [a,b] \to \mathbb{R}$ be continuous and let $h\colon [a,b] \to \mathbb{R}$ be monotone. Then there exists $c \in (a,b)$ such that*

$$\int_a^b \phi(x)h(x)\,\mathrm{d}x = h(b^-)\int_c^b \phi(x)\,\mathrm{d}x + h(a^+)\int_a^c \phi(x)\,\mathrm{d}x.$$

With this in hand, we now proceed to the proof of Jordan's criterion:

Proof of theorem 1.6. As $D_N(y)$ is even, we can rewrite

$$S_N f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)D_N(y)\,\mathrm{d}y = \frac{1}{2\pi}\int_0^{\pi} (f(x-y) + f(x+y))D_N(y)\,\mathrm{d}y.$$

As every BV function f is the difference of two monotonic functions, it suffices to show that

$$\frac{1}{2\pi}\int_0^{\pi} g(y)D_N(y)\,\mathrm{d}y \to \frac{g(0^+)}{2}$$

as $N \to \infty$, where g is monotone, since then we can take $g(y) = f(x+y)$ and $g(y) = f(x-y)$ to complete the result. Define $\tilde{g}(y) = g(y) - g(0^+)$; notice that

$$\frac{1}{2\pi}\int_0^\pi \tilde{g}(y)D_N(y)\,\mathrm{d}y \to \frac{\tilde{g}(0^+)}{2} = 0$$

if and only if

$$\frac{1}{2\pi}\int_0^\pi (g(y) - g(0^+))D_N(y)\,\mathrm{d}y \to 0$$

if and only if

$$\frac{1}{2\pi}\int_0^\pi g(y)D_N(y)\,\mathrm{d}y \to \frac{g(0^+)}{2},$$

as $\frac{1}{2\pi}\int_0^\pi D_N(y) = \frac{1}{2}$. Thus, without loss of generality, suppose that $g(0^+) = 0$ and that g is monotone increasing. We now use lemma 1.7 to prove that

$$\frac{1}{2\pi}\int_0^\pi g(y)D_N(y)\,\mathrm{d}y \to 0$$

as $N \to \infty$. As $g(0^+) = 0$, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $g(x) < \varepsilon$ whenever $x < \delta$. Then

$$\frac{1}{2\pi}\int_0^\pi g(y)D_N(y)\,\mathrm{d}y = \underbrace{\frac{1}{2\pi}\int_0^\delta g(y)D_N(y)\,\mathrm{d}y}_{=:I_1} + \underbrace{\frac{1}{2\pi}\int_\delta^\pi g(y)D_N(y)\,\mathrm{d}y}_{=:I_2}.$$

Now

$$I_2 = \frac{1}{2\pi}\int_\delta^\pi \frac{g(y)}{\sin(y/2)}\sin(N+\tfrac{1}{2})y\,\mathrm{d}y = \frac{1}{2\pi}\int_0^\pi \underbrace{\frac{g(y)}{\sin(y/2)}\chi_{[\delta,\pi]}}_{\in L^1}\sin(N+\tfrac{1}{2})y\,\mathrm{d}y \to 0$$

as $N \to \infty$, by the Riemann–Lebesgue lemma. By the mean value formula above, taking $h = g$ and $\phi = D_N$, we have that there exists $C \in (0,\delta)$ such that

$$\begin{aligned}
I_1 = \frac{1}{2\pi}\int_0^\delta g(y)D_N(y)\,\mathrm{d}y &= g(\delta^-)\int_C^\delta D_N(y)\,\mathrm{d}y \\
&\le \frac{1}{2\pi}\varepsilon \sup_{c,\delta,N}\int_C^\delta D_N(y)\,\mathrm{d}y.
\end{aligned}$$

As long as the sup is finite, we can send $\varepsilon \to 0$ and we are done, so:

$$\begin{aligned}
\left|\int_C^\delta D_N(y)\,\mathrm{d}y\right| &= \left|\int_C^\delta \sin(N+\tfrac{1}{2})y\left[\frac{1}{\sin(y/2)} - \frac{1}{y/2}\right]\mathrm{d}y\right| + \left|\int_C^\delta \frac{\sin(N+\frac{1}{2})y}{y/2}\,\mathrm{d}y\right| \\
&\le K_1 + 2\sup_{M>0}\left|\int_0^M \frac{\sin(y)}{y}\,\mathrm{d}y\right|,
\end{aligned}$$

which is bounded independent of c, δ and N. This completes the proof. $\square$

Dini's theorem and Jordan's criterion may lead one to think that getting a Fourier series to converge is relatively easy. Unfortunately, it is not. du Bois Reymond showed in 1873 that even if f is continuous, it is possible for the Fourier series to diverge at a point.

Theorem 1.8 (du Bois Reymond, 1873)**.** *There exists a continuous function $f\colon \mathbb{T} \to \mathbb{R}$ for which $S_N f(x)$ diverges for at least one x.*

Taking such an f from this theorem, and assuming (without loss of generality) that $S_N f(0)$ diverges, we can, by enumerating the rational numbers as $(r_n)_{n\in\mathbb{N}}$, construct a function $g(x) = \sum_{n=1}^{\infty} \frac{f(x-r_n)}{2^n}$ whose Fourier series diverges at every rational point.

To prove du Bois Reymond's theorem, we need the uniform boundedness principle from functional analysis:

Lemma 1.9 (Uniform boundedness principle)**.** *If X is a normed vector space and Y is a Banach space, and $T_\alpha\colon X \to Y$ is a collection of bounded linear maps for each $\alpha \in \Lambda$ (where Λ is any index set, not necessarily countable), then either*

(i) $\sup_{\alpha\in\Lambda} \|T_\alpha\|_{\mathrm{op}} < \infty$*, or*

(ii) there exists $x \in X$ such that $\sup_{\alpha\in\Lambda} \|T_\alpha x\|_Y = +\infty$.

Proof of theorem 1.8. Let $X = C(\mathbb{T})$, $Y = \mathbb{C}$, and consider the maps $T_N\colon C(\mathbb{T}) \to \mathbb{C}$ for $N \in \mathbb{N}$ given by

$$T_N f := S_N f(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) D_N(y)\,\mathrm{d}y$$

(where we have used the fact that D_N is even). As $D_N(y) = \frac{\sin(N+\frac{1}{2})y}{\sin y/2}$ has finitely many zeros, so

$$g(y) := \operatorname{sgn}(D_N(y))$$

is a measurable (but not continuous) function. We would like to consider

$$L_N := T_N(g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(y)|\,\mathrm{d}y;$$

indeed, the definition of T_N makes sense for any $f \in L^1$, whence we have that

$$|T_N f| = |S_N f(0)| \le L_N \|f\|_{L^\infty}$$

and hence that $\|T_N\|_{\mathrm{op}} \le L_N$. (We have bounded the operator norm of T_N as an operator from $L^1(\mathbb{T}) \to \mathbb{C}$, and used the fact that restriction never increases the norm.) Clearly by using g we can see that the operator norm of T_N as an operator from $L^1(\mathbb{T}) \to \mathbb{C}$ should be exactly L_N; as g is not continuous, we use the fact that continuous functions are dense to see that, given $\varepsilon > 0$ there exists an $h \in C^0(\mathbb{T})$ such that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} h(y) D_N(y)\,\mathrm{d}y \ge L_N - \varepsilon.$$

Hence the operator norm of $T_N\colon C(\mathbb{T}) \to \mathbb{C}$ is $\|T_N\|_{\mathrm{op}} = L_N$.

By the uniform boundedness principle, either $L_N \leq K$ for some $K < \infty$ and all $n \in \mathbb{N}$, or there exists $f \in C(\mathbb{T})$ such that $\limsup_{N\to\infty} |T_N f| = +\infty$. We wish to exclude the first possibility to show that there is a function f such that $S_N f(0)$ diverges; to do so, we prove that $L_N \to +\infty$, by showing that

$$L_N = \frac{1}{2\pi}\int_{-\pi}^{\pi} |D_N(y)|\,\mathrm{d}y = \frac{4}{\pi^2}\log N + O(1). \tag{1}$$

To see this, we compute:

$$\begin{aligned}
L_N &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\sin(N+\frac{1}{2})y}{\sin y/2}\right|\mathrm{d}y \\
&= \frac{1}{\pi}\int_0^{\pi}\left|\frac{\sin(N+\frac{1}{2})y}{\sin y/2}\right|\mathrm{d}y && \text{as } D_N \text{ is even} \\
&= \frac{1}{\pi}\int_0^{\pi}\frac{|\sin(N+\frac{1}{2})y|}{\sin y/2}\,\mathrm{d}y && \text{as } \sin \geq 0 \text{ on } [0,\pi/2] \\
&= \frac{1}{\pi}\int_0^{\pi}|\sin(N+\tfrac{1}{2})y|\left[\frac{1}{\sin y/2} - \frac{1}{y/2} + \frac{1}{y/2}\right]\mathrm{d}y \\
&= \frac{1}{\pi}\int_0^{\pi}|\sin(N+\tfrac{1}{2})y|\left[\frac{1}{\sin y/2} - \frac{1}{y/2}\right]\mathrm{d}y + \frac{2}{\pi}\int_0^{\pi}\frac{|\sin(N+\frac{1}{2})y|}{y}\,\mathrm{d}y.
\end{aligned}$$

As $y \mapsto \left[\frac{1}{\sin y/2} - \frac{1}{y/2}\right]$ is bounded, the first integral is $O(1)$. Consider the second integral under the change of variables $(N+\frac{1}{2})y = \pi z$:

$$\begin{aligned}
L_N &= \frac{2}{\pi}\int_0^{\pi}\frac{|\sin(N+\frac{1}{2})y|}{y}\,\mathrm{d}y + O(1) \\
&= 2\int_0^{N+\frac{1}{2}}\frac{|\sin \pi z|}{\pi z}\,\mathrm{d}z + O(1) \\
&= 2\sum_{k=0}^{N}\int_k^{k+1}\frac{|\sin \pi z|}{\pi z}\,\mathrm{d}z + O(1) \\
&= 2\sum_{k=0}^{N}\int_0^{1}\frac{|\sin \pi z|}{\pi(z+k)}\,\mathrm{d}z + O(1) \\
&= \frac{2}{\pi}\int_0^1 |\sin \pi z|\sum_{k=0}^{N}\frac{1}{z+k}\,\mathrm{d}z + O(1) \\
&= \frac{2}{\pi}\log N \underbrace{\int_0^1 |\sin \pi z|\,\mathrm{d}z}_{2/\pi} + O(1) \\
&= \frac{4}{\pi^2}\log N + O(1).
\end{aligned}$$

□

The moral of this theorem is that pointwise convergence is, quite simply, too much to ask. Kolmogorov showed just how much it is to ask that the Fourier series of a function converge pointwise:

Theorem 1.10 (Kolmogorov, 1926)**.** *There exists $f \in L^1(\mathbb{T})$ such that $S_N f(x)$ diverges at* every *point $x \in \mathbb{T}$.*

The proof of this theorem is beyond the scope of the course. The existence of a function in $L^1(\mathbb{T})$ whose Fourier series diverges almost everywhere is demonstrated in section 3.4.2 of [GraCl].

Having seen that pointwise convergence is, ultimately, fruitless in many cases, we move on to a different flavour of convergence results. Given a function f, form the partial sums of the Fourier series $S_N f$.

1. Does $S_N f \to f$ in the L^p norm?
2. Does $S_N f \to f$ almost everywhere?

For the second question, Carleson and Hunt showed that this Kolmogorov's example of a function whose Fourier series diverges is largely due to the nature of L^1, and that considering L^p for $1 < p < \infty$ actually gains us *almost everywhere* pointwise convergence:

Theorem 1.11 (Carleson, 1965)**.** *If $f \in L^2(\mathbb{T})$, then $S_N f(x)$ converges to $f(x)$ for almost every $x \in \mathbb{T}$.*

Carleson's theorem, which won him the Abel prize, was extended by Hunt a few years later:

Theorem 1.12 (Hunt, 1967)**.** *Let $1 < p < \infty$. If $f \in L^p(\mathbb{T})$, then $S_N f(x)$ converges to $f(x)$ for almost every $x \in \mathbb{T}$.*

Again, the proof of the Carleson–Hunt theorem is beyond the scope of the course: in fact, it occupies the whole of chapter 11 of [GraMo], and is *not* for the faint of heart.

With regard to the first question, the main theorem on convergence in the L^p norm — which will take some time to build up to — is the following:

Theorem 1.13. *Let $1 < p < \infty$. If $f \in L^p(\mathbb{T})$, then $S_N f \to f$ in the L^p norm; that is, $\|S_N f - f\|_{L^p} \to 0$.*

The theorem is easy enough to prove in L^2 as it is a Hilbert space; the real meat of the theorem is for $p \neq 2$, where L^p is a Banach space but not a Hilbert space. We will prove the result using the following key step:

Proposition 1.14. *Let $1 < p < \infty$. The following are equivalent:*

- *for all $f \in L^p(\mathbb{T})$, $S_N f \to f$ in the L^p norm;*
- *there exists a constant c_p such that, for all $f \in L^p(\mathbb{T})$ and all $N \in \mathbb{N}$,*

$$\|S_N f\|_{L^p} \leq c_p \|f\|_{L^p}.$$

Proof. First, we show that, if $S_N f \to f$ in the L^p norm for all $f \in L^p(\mathbb{T})$, then such a constant c_p exists. Consider $S_N \colon L^p(\mathbb{T}) \to \mathbb{C}$ as operators for each $N \in \mathbb{N}$. Each S_N is a bounded linear operator, so by the uniform boundedness principle, either

(i) $\sup_{N\in\mathbb{N}} \|S_N\|_{\mathrm{op}} < \infty$, or

(ii) there exists $f \in L^p$ such that $\limsup_{N\to\infty} \|S_N f\|_{L^p} = +\infty$.

We show that (i) holds. Given $\varepsilon > 0$ and $f \in L^p(\mathbb{T})$, pick N large enough so that $\|S_N f - f\|_{L^p} < \varepsilon$. Then for such large enough N,

$$\|S_N f\|_{L^p} \leq \|S_N f - f\|_{L^p} + \|f\|_{L^p} \leq \varepsilon + \|f\|_{L^p}.$$

This holds for all $\varepsilon > 0$, and neither $\|S_N f\|_{L^p}$ nor $\|f\|_{L^p}$ depend on ε, so we have $\|S_N f\|_{L^p} \leq \|f\|_{L^p}$ for all f and all N large enough. Hence

$$c_p := \sup_{N\in\mathbb{N}} \|S_N\|_{\mathrm{op}}$$

exists and is finite.

For the converse — that is, showing that the existence of such a constant c_p guarantees that $S_N f \to f$ in the L^p norm for all $f \in L^p$ — we first note that trigonometric polynomials are dense in L^p whenever $1 < p < \infty$: that is, given $\varepsilon > 0$ and $f \in L^p$, there exists a function $g\colon \mathbb{T} \to \mathbb{R}$ of the form

$$g(x) = \sum_{k=-M}^{M} a_k e^{ikx}$$

such that $\|f - g\|_{L^p} < \varepsilon$. So, fix $\varepsilon > 0$ and $f \in L^p$, and let g be a trigonometric polynomial such that $\|f - g\|_{L^p} < \frac{\varepsilon}{1+c_p}$. Then whenever $N > \deg(g)$, we have that $S_N g = g$, so that

$$\begin{aligned}
\|S_N f - f\|_{L^p} &\leq \|S_N f - S_N g\|_{L^p} + \|S_N g - g\|_{L^p} + \|g - f\|_{L^p} \\
&= \|S_N(f - g)\|_{L^p} + 0 + \|f - g\|_{L^p} \\
&\leq (1 + c_p)\|f - g\|_{L^p} \\
&< (1 + c_p)\frac{\varepsilon}{1 + c_p} = \varepsilon.
\end{aligned}$$

Hence $S_N f \to f$ in the L^p norm, for any $f \in L^p$. □

To see that trigonometric polynomials are dense in $L^p(\mathbb{T})$, recall that if $f \in C^2(\mathbb{T})$ then $S_N f \to f$ uniformly; that is, given $\varepsilon > 0$, there exists N such that for $n \geq N$,

$$|S_n f(x) - f(x)| < \frac{\varepsilon}{(2\pi)^{1/p}}$$

for all $x \in \mathbb{T}$. Observe that $S_N f$ is a trigonometric polynomial, and that

$$\|S_N f - f\|_{L^p} = \left(\int_{-\pi}^{\pi} |S_N f(x) - f(x)|^p \,\mathrm{d}x\right)^{1/p} \leq \frac{\varepsilon}{(2\pi)^{1/p}} \cdot (2\pi)^{1/p} = \varepsilon,$$

so the trigonometric polynomials are dense in $C^2(\mathbb{T})$, and $C^2(\mathbb{T})$ is dense in $L^p(\mathbb{T})$.

1.3 Good Kernels and PDEs

We saw that the convergence of $S_N f$ is closely related to the properties of the Dirichlet kernel D_N, by the equation

$$S_N f(x) = \frac{1}{2\pi}(f * D_N)(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y) D_N(x-y)\,\mathrm{d}y.$$

We exploited the fact that $\frac{1}{2\pi}\int_{-\pi}^{\pi} D_N(y)\,\mathrm{d}y = 1$ for all N; we also exploited the fact that for *fixed* $\delta > 0$,

$$\int_{\delta<|y|<\pi} D_N(y) \to 0$$

as $N \to \infty$. Indeed, most of the proofs relied on only these two facts. Unfortunately, we also saw that

$$\int_{-\pi}^{\pi} |D_N(y)|\,\mathrm{d}y = \frac{4}{\pi^2}\log N + O(1) \to \infty$$

as $N \to \infty$; if these integrals had been bounded uniformly in N some of the proofs would have been much easier.

We now consider other modes of convergence, and other ways of summing Fourier series: it turns out that the partial sums can also be expressed as the convolution of f with some kernel K_N. The other modes of convergence we will investigate, however, will have much nicer convergence properties than standard summation via convolution with the Dirichlet kernels. We generalise and define a *good kernel* as follows:

Definition 1.15 (Good kernel)**.** *Let* $K_n \colon \mathbb{T} \to \mathbb{R}$ *for* $n \in \mathbb{N}$*.* (K_n) *is a* family of good kernels *if*

(i) $\displaystyle\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(x)\,\mathrm{d}x = 1$*;*

(ii) there exists a constant $K > 0$ *such that, for all* $n \in \mathbb{N}$*,* $\displaystyle\int_{-\pi}^{\pi} |K_n(x)|\,\mathrm{d}x \le K$*; and*

(iii) for every $\delta > 0$*,* $\displaystyle\int_{\delta<|x|<\pi} |K_n(x)|\,\mathrm{d}x \to 0$ *as* $n \to \infty$*.*

So D_N satisfies (i) and (iii), but not (ii), and thus is not a good kernel. The difference this makes is readily apparent: with property (ii), we can show that, in contrast to the theorem of du Bois Reymond, if K_n is a family of good kernels, then $K_n * f(x) \to f(x)$ for every point of continuity of f:

Theorem 1.16. *Let* (K_n) *be a* family of good kernels*, and let* $f \in L^1(\mathbb{T}) \cap L^\infty(\mathbb{T})$*. If* f *is continuous at* $x \in \mathbb{T}$*, then*

$$\left|\frac{1}{2\pi}(K_n * f)(x) - f(x)\right| \to 0$$

as $n \to \infty$*. In particular, if* $f \in C^0(\mathbb{T})$*, then the convergence is uniform; that is, for all* $\varepsilon > 0$ *there exists* N *such that for all* $x \in \mathbb{T}$ *and* $n \ge N$*,*

$$\left|\frac{1}{2\pi}(K_n * f)(x) - f(x)\right| < \varepsilon.$$

Proof. Let x be a point of continuity of f, and fix $\varepsilon > 0$. By property (ii), there exists K such that for all $n \in \mathbb{N}$,

$$\int_{-\pi}^{\pi} |K_n(x)| \, \mathrm{d}x \leq K.$$

Take $\delta > 0$ such that whenever $|y| < \delta$, we have $|f(x-y) - f(x)| < \frac{\pi\varepsilon}{K}$. Given this δ, using property (iii) choose N such that, for $n \geq N$,

$$\int_{\delta<|y|<\pi} |K_n(y)| \, \mathrm{d}y \leq \frac{\pi\varepsilon}{2\|f\|_\infty}.$$

Then, for $n \geq N$,

$$\begin{aligned}
&\left|\frac{1}{2\pi}(K_n * f)(x) - f(x)\right| \\
&\quad = \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)\,(f(x-y) - f(x))\,\mathrm{d}y\right| && \text{by property (i)} \\
&\quad \leq \left|\frac{1}{2\pi}\int_{-\delta}^{\delta} K_n(y)\,(f(x-y) - f(x))\,\mathrm{d}y\right| + \left|\frac{1}{2\pi}\int_{\delta\leq|y|\leq\pi} K_n(y)\,(f(x-y) - f(x))\,\mathrm{d}y\right| \\
&\quad \leq \frac{1}{2\pi}\int_{-\delta}^{\delta} |K_n(y)|\,\underbrace{|f(x-y) - f(x)|}_{\leq\frac{\pi\varepsilon}{K}}\,\mathrm{d}y + \frac{1}{2\pi}\int_{\delta\leq|y|\leq\pi} |K_n(y)|\,\underbrace{|f(x-y) - f(x)|}_{\leq 2\|f\|_\infty}\,\mathrm{d}y \\
&\quad \leq \frac{\varepsilon}{2K}\int_{-\delta}^{\delta} |K_n(y)|\,\mathrm{d}y + \frac{\|f\|_\infty}{\pi}\int_{\delta<|y|<\pi} |K_n(y)|\,\mathrm{d}y \\
&\quad \leq \frac{\varepsilon}{2K}\underbrace{\int_{-\pi}^{\pi} |K_n(y)|\,\mathrm{d}y}_{\leq K \text{ for all } n} + \frac{\|f\|_\infty}{\pi}\cdot\frac{\pi\varepsilon}{2\|f\|_\infty} \\
&\quad \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Finally, note that if $f \in C^0(\mathbb{T})$, then we may choose δ independently of x, and thus we may choose N independently of x, and hence the convergence is uniform in x. □

1.4 Cesàro Summation and the Fejér Kernel

Let $(a_n)_{n=0}^{\infty}$ be a sequence, and consider the n^{th} partial sum $s_n = a_0 + a_1 + \cdots + a_n$. Does (s_n) converge? That question is related to the convergence of

$$\sigma_n = \frac{s_0 + s_1 + \cdots + s_{n-1}}{n}.$$

If $s_n \to s$, then $\sigma_n \to s$ as well. However, sometimes σ_n will converge when s_n does not. Let $a_n = (-1)^n$; then

$$s_n = \underbrace{1 - 1 + 1 - 1 + \ldots}_{n \text{ times}},$$

so $s_0 = 1$, $s_1 = 0$, $s_2 = 1$, $s_3 = 0$, and so on. So s_n does not converge. However, $\sigma_n \to \frac{1}{2}$. We call σ_n the n^{th} *Cesàro mean* of s_n. If σ_n converges to σ, but s_n does not converge, we say that $s_n \to \sigma$ in the Cesàro sense.

To apply this to Fourier series, given a function $f\colon \mathbb{T} \to \mathbb{R}$, with partial Fourier sums $S_n f(x)$, define

$$\sigma_n f(x) := \frac{S_0 f(x) + S_1 f(x) + \cdots + S_{n-1} f(x)}{n}.$$

We wish to express $\sigma_n f(x)$ as the convolution of f with some kernel, so we compute:

$$\begin{aligned}
\sigma_n f(x) &= \frac{S_0 f(x) + S_1 f(x) + \cdots + S_{n-1} f(x)}{n} \\
&= \frac{1}{n}\frac{1}{2\pi}\left[\int_{-\pi}^{\pi} f(y) D_0(x-y)\,\mathrm{d}y + \cdots + \int_{-\pi}^{\pi} f(y) D_{n-1}(x-y)\,\mathrm{d}y\right] \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(y)\left[\frac{1}{n}\sum_{k=0}^{n-1} D_k(x-y)\right]\,\mathrm{d}y
\end{aligned}$$

So we define the *Fejér kernel* as

$$F_n(t) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(t);$$

by the above, we have that

$$\sigma_n f(x) = \frac{1}{2\pi}(F_n * f)(x).$$

We now express the Fejér kernel in a more convenient form. First, note that

$$\begin{aligned}
\cos(kt) - \cos(k+1)t &\equiv \cos((k+\tfrac{1}{2})t - \tfrac{1}{2}t) - \cos((k+\tfrac{1}{2})t + \tfrac{1}{2}t) \\
&\equiv \big(\cos(k+\tfrac{1}{2})t\cos(t/2) + \sin(k+\tfrac{1}{2})t\sin(t/2)\big) \\
&\quad - \big(\cos(k+\tfrac{1}{2})t\cos(t/2) - \sin(k+\tfrac{1}{2})t\sin(t/2)\big) \\
&\equiv 2\sin(k+\tfrac{1}{2})t\sin t/2.
\end{aligned}$$

Using this, and the identity $\sin^2(nt/2) \equiv \frac{1-\cos(nt)}{2}$, we obtain:

$$\begin{aligned}
F_n(t) &= \frac{1}{n}\sum_{k=0}^{n-1} D_k(t) \\
&= \frac{1}{n}\sum_{k=0}^{n-1}\frac{\sin(k+\frac{1}{2})t}{\sin t/2} \\
&= \frac{1}{n\sin^2 t/2}\sum_{k=0}^{n-1}\sin(k+\tfrac{1}{2})t\sin t/2 \\
&= \frac{1}{n\sin^2 t/2}\left(\frac{1}{2}\sum_{k=0}^{n-1}(\cos(kt) - \cos(k+1)t)\right) \\
&= \frac{1}{n\sin^2 t/2}\cdot\frac{1-\cos(nt)}{2} \\
&= \frac{\sin^2(nt/2)}{n\sin^2(t/2)}.
\end{aligned}$$

Having expressed the Fejér kernel in closed form, we now show that it forms a family of good kernels:

Theorem 1.17. *The Fejér kernel* $F_n(t) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(t)$ *is a family of good kernels. Hence, if* $f \in L^1(\mathbb{T}) \cap L^\infty(\mathbb{T})$*, and* f *is continuous at* $x \in \mathbb{T}$*, then* $\sigma_n f(x) \to f(x)$*.*

Proof. It suffices to check that the three properties of definition 1.15 hold:

(i) We observe that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} F_n(x)\,\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{n}\sum_{k=0}^{n-1} D_k(x) = \frac{1}{2\pi}\frac{1}{n}\sum_{k=0}^{n-1}\underbrace{\int_{-\pi}^{\pi} D_k(x)}_{=2\pi} = 1.$$

(ii) As $F_n(x) \geq 0$ for all $x \in \mathbb{T}$ and all $n \in \mathbb{N}$, we see that $|F_n(x)| = F_n(x)$, so that $\int_{-\pi}^{\pi} |F_n(x)|\,\mathrm{d}x = 2\pi$ for all $n \in \mathbb{N}$.

(iii) Fix $\delta > 0$. Whenever $\delta \leq |x| \leq \pi$, we have that $\frac{1}{\sin^2(x/2)} \leq M_\delta$ for some constant M_δ which depends on δ. Thus

$$|F_n(x)| = \frac{1}{n}\left|\frac{\sin^2(Nt/2)}{\sin^2(t/2)}\right| \leq \frac{M_\delta}{n}$$

for all x such that $\delta \leq |x| \leq \pi$, and hence

$$\int_{\delta<|x|<\pi} |F_n(x)|\,\mathrm{d}x \leq \frac{2\pi M_\delta}{n} \to 0$$

as $n \to \infty$, as required. □

The theorem implies that, if x is a point of continuity of f, then there exists a sequence $\sigma_n f(x)$ of trigonometric polynomials which converge to f. (Hence trigonometric polynomials are dense in $C^0(\mathbb{T})$, and hence they are dense in $L^p(\mathbb{T})$.)

The difference between S_n and σ_n can be summarised as follows:

- Going from $S_n f$ to $S_{n+1} f$, you do *not* change the first $2n+1$ Fourier coefficients: the first $2n+1$ Fourier coefficients of $S_{n+1} f$ are exactly the same as those of $S_n f$.
- Going from $\sigma_n f$ to $\sigma_{n+1} f$, you must recompute *every* Fourier coefficient!

Corollary 1.18. *Let* $f \in L^1(\mathbb{T})$*. Suppose that* $\hat{f}(n) = 0$ *for all* $n \in \mathbb{Z}$*. Then* $f(x) = 0$ *for all points of continuity of* f*; in particular,* $f(x) = 0$ *for almost every* $x \in \mathbb{T}$*.*

As an application of the Fejér kernel, let us exhibit an example of a function which is continuous but nowhere differentiable; that is, a continuous function $f\colon \mathbb{T} \to \mathbb{R}$ such that $f'(x)$ does not exist for any $x \in \mathbb{T}$. To do so, we prove a theorem showing that, given a function with Fourier coefficients of a particular form which is differentiable at some point, said Fourier coefficients must satisfy an estimate. We then exhibit a function which does not satisfy any such estimate, and which thus cannot be differentiable at any point.

Theorem 1.19. *Let $g \in C^0(\mathbb{T})$ be periodic and continuous such that*

$$\hat{g}(n) = \begin{cases} a_m^\pm & \text{if } n = \pm 2^m, \text{ for some } m \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

If g is differentiable at x_0, then there exists a constant C such that

$$|a_m^\pm| \leq \frac{Cm}{2^m}$$

whenever $m \neq 0$.

Proof. Without loss of generality we may assume that $x_0 = 0$, since otherwise we may put $h(x) = g(x - x_0)$, which has

$$|\hat{h}(n)| = |\hat{g}(n)|.$$

Furthermore, without loss of generality we may assume that $g(0) = 0$, since otherwise we may consider $g(x) - g(0)$, which will have the same Fourier coefficients except for $n = 0$.

As g is differentiable at $x = 0$, g is locally Lipschitz around 0: that is, there exists $K_1 > 0$ and $\delta > 0$ such that whenever $|x| < \delta$, we have

$$|g(x)| \leq K_1|x|.$$

As g is continuous, so is $\frac{g(x)}{|x|}$ for $\delta \leq |x| \leq \pi$; set

$$K_2 := \sup\left\{\frac{g(x)}{|x|} : \delta \leq |x| \leq \pi\right\}$$

(which is finite as $[-\pi, -\delta] \cup [\delta, \pi]$ is compact). Then, for $K := \max\{K_1, K_2\}$, we have that

$$|g(x)| \leq K|x|$$

for all $x \in [-\pi, \pi]$. Notice that, for $x \in [-\pi, \pi]$, we have $|\sin(x/2)| \geq \frac{1}{\pi}|x|$, so

$$\frac{|g(x)|}{|\sin^2(x/2)|} \leq K\pi^2|x|$$

Recall that $e_m(x) := e^{imx}$. We claim that

$$a_m^+ = \frac{1}{2\pi}\langle g, e_{2^m}F_M\rangle$$

for $M = 2^{m-1} - 1$. To see this, consider that

$$F_M(x) = \frac{1}{M}\sum_{k=0}^{M-1}\sum_{j=-k}^{k} e^{ijx} = \frac{1}{M}\sum_{k=-(M-1)}^{M-1}(M - |k|)e^{ikx}.$$

So F_M is a sum of exponentials between $-(2^{m-1} - 1)$ and $(2^{m-1} - 1)$; note that the coefficient of e^{ikx} when $k = 0$ is $\frac{M-|k|}{M} = 1$. Hence $e_{2^m}F_M$ is a sum of exponentials between $2^m - (2^{m-1} - 1) \geq 2^{m-1} - 1$ and $2^m + (2^{m-1} - 1) \leq 2^{m+1} - 1$, and the coefficient in front of $e^{i2^m x}$ is 1. As the Fourier coefficients $\hat{g}(n)$ of g are 0 unless n is a power of 2, we see that

$$\langle g, e_{2^m} F_M \rangle = \langle g, e_{2^m} \rangle = 2\pi \hat{g}(2^m) = 2\pi a_m^+,$$

as required. We now use this to estimate the value of $|a_m^+|$:

$$\begin{aligned}
|a_m^+| &= \frac{1}{2\pi} \left| \langle g, e_{2^m} F_M \rangle \right| \\
&= \frac{1}{2\pi} \left| \int_{-\pi}^{\pi} g(x) \overline{e_{2^m}(x) F_M(x)} \, \mathrm{d}x \right| \\
&= \frac{1}{2\pi} \left| \int_{-\pi}^{\pi} g(x) e^{-i2^m x} \frac{\sin^2(Mx/2)}{M \sin^2(x/2)} \, \mathrm{d}x \right| \\
&\leq \frac{1}{2\pi M} \int_{-\pi}^{\pi} |g(x)| \frac{\sin^2(Mx/2)}{\sin^2(x/2)} \, \mathrm{d}x \\
&\leq \frac{K\pi^2}{2\pi M} \int_{-\pi}^{\pi} \frac{\sin^2(Mx/2)}{|x|} \, \mathrm{d}x \\
&= \frac{K\pi}{M} \int_{0}^{\pi} \frac{\sin^2(Mx/2)}{|x|} \, \mathrm{d}x \\
&= \frac{K\pi}{M} \int_{0}^{1/M} \frac{\sin^2(Mx/2)}{|x|} \, \mathrm{d}x + \frac{K\pi}{M} \int_{1/M}^{\pi} \frac{\sin^2(Mx/2)}{|x|} \, \mathrm{d}x \\
&\leq \frac{K\pi}{4M} \int_{0}^{1/M} \frac{M^2 x^2}{|x|} \, \mathrm{d}x + \frac{K\pi}{M} \int_{1/M}^{\pi} \frac{1}{|x|} \, \mathrm{d}x \\
&\leq \frac{K\pi}{8M} + \frac{K\pi}{M} (\log M + \log \pi) \\
&= \frac{K\pi}{M} \left(\frac{1}{8} + \log \pi + \log M \right) \\
&\leq \frac{C}{2^{m-1} - 1} (2 + \log(2^{m-1} - 1)) \\
&\leq \frac{Cm}{2^m}
\end{aligned}$$

for some constant C. □

We now define

$$f(x) = \sum_{n=1}^{\infty} a^n \cos(2^n x)$$

for $\frac{1}{2} < a < 1$. Noting that

$$\cos(2^n x) = \frac{e^{i2^n x} + e^{-i2^n x}}{2},$$

we may write

$$f(x) = \sum_{n \in \mathbb{Z} \setminus \{0\}} \frac{1}{2} a^{|n|} e^{i2^n x}.$$

Thus

$$\hat{f}(n) = \begin{cases} \frac{1}{2}a^{|m|} & \text{if } n = \pm 2^m \text{ for some } m \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

If f is differentiable, then the coefficients must satisfy an estimate of the form $\frac{1}{2}a^{|m|} \leq \frac{Cm}{2^m}$, or, equivalently, $\frac{2^{m-1}}{m}a^{|m|} \leq C$. We show that no such estimate can hold: observe that

$$\frac{2^{m-1}}{m}a^{|m|} = \frac{1}{2m}\left(\frac{a}{1/2}\right)^m \to +\infty$$

as $m \to \infty$, since $\frac{a}{1/2} > 1$. Since no such estimate can hold, f cannot be differentiable at any point.

1.5 Abel Summation and the Poisson Kernel

Given a series $\sum_{k=0}^{\infty} a_k$, which may or may not converge, but has $|a_k| \leq M$ for all $k \in \mathbb{N}$, define

$$A(r) = \sum_{k=0}^{\infty} a_k r^k.$$

As $|a_k| \leq M$, $A(r)$ is well defined for $|r| < 1$. If $A(1^-) := \lim_{r \to 1^-} A(r)$ exists, we denote it by $A(1)$ and say that

$$\sum_{k=0}^{\infty} a_k = A(1)$$

in the Abel sense. (Of course, if $\sum_{k=0}^{\infty} a_k$ converges, then $A(1)$ always exists and equals $\sum_{k=0}^{\infty} a_k$ in the usual sense.)

For example, let us consider $\sum_{k=0}^{\infty}(-1)^k(k+1) = 1 - 2 + 3 - 4 + \ldots$. Technically this does not fit in to the above definition, but the power series $\sum_{k=0}^{\infty}(-1)^k(k+1)r^k$ converges for $|r| < 1$; notice that

$$A(r) = \sum_{k=0}^{\infty}(-1)^k(k+1)r^k = \frac{1}{(1+r)^2}.$$

So

$$1 - 2 + 3 - 4 + \cdots = \frac{1}{4}$$

in the Abel sense.

Once again, we apply this to Fourier series. Given $f \in L^1(\mathbb{T})$, for $0 \leq r < 1$, we define

$$A_r f(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n) r^{|n|} e^{in\theta}.$$

Once again we want to show that $A_r f(\theta) = \frac{1}{2\pi}(f * P_r)(\theta)$ for some kernel P_r:

$$\begin{aligned}
A_r f(\theta) &= \sum_{n=-\infty}^{\infty} \hat{f}(n) r^{|n|} e^{in\theta} \\
&= \sum_{n=-\infty}^{\infty} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} \,\mathrm{d}x \, r^{|n|} e^{in\theta} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \sum_{n=-\infty}^{\infty} r^{|n|} e^{in(\theta - x)} \,\mathrm{d}x \qquad \text{by the Dominated Convergence Theorem} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) P_r(\theta - x) \,\mathrm{d}x
\end{aligned}$$

where $P_r(\theta) := \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta}$ is called the *Poisson kernel*. Again, we compute P_r in closed form:

$$\begin{aligned}
P_r(\theta) &= \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta} \\
&= \sum_{n=0}^{\infty} r^{|n|} e^{in\theta} + \sum_{n=1}^{\infty} r^{|n|} e^{-in\theta} \\
&= \sum_{n=0}^{\infty} (re^{i\theta})^n + \sum_{n=1}^{\infty} (re^{-i\theta})^n \\
&= \sum_{n=0}^{\infty} \omega^n + \bar{\omega} \sum_{n=1}^{\infty} \bar{\omega}^n && \text{for } \omega := re^{i\theta} \\
&= \frac{1}{1-\omega} + \frac{\bar{\omega}}{1-\bar{\omega}} \\
&= \frac{(1-\bar{\omega}) + \bar{\omega}(1-\omega)}{(1-\omega)(1-\bar{\omega})} \\
&= \frac{1-|\omega|^2}{|1-\omega|^2} && \text{since } \omega\bar{\omega} = |\omega|^2 \\
&= \frac{1-r^2}{|1-re^{it}|^2} \\
&= \frac{1-r^2}{1-2r\cos(\theta)+r^2}.
\end{aligned}$$

Theorem 1.20. *The Poisson kernel $P_r(\theta) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta}$ is a family of good kernels. Hence, if $f \in L^1(\mathbb{T}) \cap L^\infty(\mathbb{T})$, and f is continuous at $x \in \mathbb{T}$, then $A_r f(x) \to f(x)$ as $r \to 1$.*

(Technically, theorem 1.16 applies as $n \to \infty$; however, it is easy to see that the same theorem will hold for convergence as $r \to 1$.)

Proof. Again, it suffices to check that the three properties of definition 1.15 hold:

(i) We observe that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(\theta)\,\mathrm{d}\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}\sum_{n=-\infty}^{\infty} r^{|n|}e^{in\theta}\,\mathrm{d}\theta = \frac{1}{2\pi}\sum_{n=-\infty}^{\infty} r^{|n|}\int_{-\pi}^{\pi} e^{in\theta}\,\mathrm{d}\theta = 1,$$

since $\int_{-\pi}^{\pi} e^{in\theta}\,\mathrm{d}\theta = 0$ unless $n = 0$, when it equals 2π.

(ii) As $P_r(\theta) \geq 0$ for all $\theta \in \mathbb{T}$ and all $r \in [0, 1)$, we see that $|P_r(\theta)| = P_r(\theta)$, so that $\int_{-\pi}^{\pi} |P_r(\theta)|\,\mathrm{d}\theta = 2\pi$ for all $r \in [0, 1)$.

(iii) Fix $\delta > 0$. We rewrite the denominator of $P_r(\theta)$ as follows:

$$1 - 2r\cos(\theta) + r^2 = (1 - r)^2 + 2r(1 - \cos\theta).$$

Now, whenever $\frac{1}{2} \leq r < 1$, there exists $C_\delta > 0$ such that whenever $\delta < |\theta| < \pi$, we have

$$1 - 2r\cos(\theta) + r^2 \geq C_\delta > 0.$$

Thus

$$\int_{\delta<|\theta|<\pi} |P_r(\theta)|\,\mathrm{d}\theta \leq \frac{1}{C_\delta}\int_{\delta<|\theta|<\pi} 1 - r^2\,\mathrm{d}\theta \leq \frac{2\pi(1 - r^2)}{C_\delta} \to 0$$

as $r \to 1$, as required. $\square$

As an application, we consider the Laplace equation on the unit ball $B \subsetneq \mathbb{R}^2$:

$$\begin{cases} -\Delta u = 0 & \text{in } B \\ u = f & \text{on } \partial B \end{cases}$$

In polar coordinates (r, θ), the Laplacian becomes

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}.$$

Attempting a solution by separation of variables, let $u(r, \theta) = F(r)G(\theta)$; the Laplace equation then becomes

$$F''G + \frac{1}{r}F'G + \frac{1}{r^2}FG'' = 0.$$

Rearranging, we obtain that

$$\frac{r^2F'' + rF'}{F} = -\frac{G''}{G} = \lambda.$$

Since the left-hand side does not depend on θ, and the right-hand side does not depend on r, both sides must in fact depend on neither and be constant, and thus both sides equal some constant $\lambda \in \mathbb{R}$. So we obtain the coupled equations

$$\begin{cases} G'' + \lambda G = 0 \\ r^2F'' + rF' - \lambda F = 0 \end{cases}$$

We require that G is 2π-periodic, so the equation for G will have solutions if and only if $\lambda = m^2$ for some $m \in \mathbb{Z}$. The solutions are

$$G(\theta) = Ae^{im\theta} + Be^{-im\theta}.$$

The solutions to the second equation depend on whether $m = 0$ or not. If $m = 0$, then we get the two linearly independent solutions $F_1(r) = 1$ and $F_2(r) = \log r$. On the other hand, if $m \neq 0$, we get the two solutions $F_1(r) = r^m$, $F_2(r) = r^{-m}$. We only really want solutions such that $u(r, \theta)$ is bounded on B, so we consider only the solutions

$$F(r) = r^{|m|}$$

for $m \in \mathbb{Z}$. Summing over all possible solutions for $m \in \mathbb{Z}$, we thus arrive at our postulated solution, given by *Poisson's formula*:

$$u(r, \theta) = \sum_{n=-\infty}^{\infty} \alpha_n r^{|n|} e^{in\theta},$$

where the α_n are constants to be determined by the boundary conditions. If $u(r, \theta) = f(\theta)$ at the boundary ∂B, then we would like to have that

$$\lim_{r \to 1^-} u(r, \theta) = f(\theta),$$

that is,

$$\lim_{r \to 1^-} \sum_{n=-\infty}^{\infty} \alpha_n r^{|n|} e^{in\theta} = f(\theta).$$

Interchanging the limit and the summation, we see that there can only be one choice of coefficients $\alpha_n = \hat{f}(n)$. In that case, we have that

$$u(r, \theta) = (f * P_r)(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n) r^{|n|} e^{in\theta},$$

is a solution that satisfies the required boundary conditions; what's more, as P_r is a family of good kernels, we see that

$$(f * P_r)(\theta) \to f(\theta)$$

as $r \to 1^-$ for every point of continuity of f. So we have proved the following theorem:

Theorem 1.21. *Let $f \in L^1(\mathbb{T}) \cap L^\infty(T)$. The unique (rotationally invariant) solution of*

$$\begin{cases} -\Delta u = 0 & \text{in } B \\ u = f & \text{on } \partial B \end{cases}$$

is given by

$$u(r, \theta) = (f * P_r)(\theta) = \sum_{n=-\infty}^{\infty} \hat{f}(n) r^{|n|} e^{in\theta} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta - y) \frac{1 - r^2}{1 - 2r\cos y + r^2} \, \mathrm{d}y$$

and satisfies $\lim_{r \to 1^-} u(r, \theta) = f(\theta)$ for every point of continuity of f.

2 Fourier Transform

Recall that $\{\frac{1}{\sqrt{2\pi}}e^{inx}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of $L^2([-\pi,\pi])$. We generalise the definition of Fourier series to an interval $[-L/2, L/2]$ of length L by defining $e_{n,L} := \frac{1}{\sqrt{L}}e^{2\pi inx/L}$, and setting

$$\hat{f}_L(n) = \frac{1}{\sqrt{L}}\int_{-L/2}^{L/2} f_L(x)e^{-2\pi inx/L}\,\mathrm{d}x = \langle f_L, e_{n,L}\rangle$$

for some $f_L \in L^2([-L/2, L/2])$.

While Fourier series are an excellent tool for functions on a compact interval (which we can think of as being periodic on all of $\mathbb{R}$), if we have a non-periodic function on all of $\mathbb{R}$ we seemingly cannot use Fourier series. In general, let us write $g_L(\xi) = \sqrt{L}\hat{f}_L(n)$ for $\xi \in [2\pi n/L, 2\pi(n+1)/L]$; observe that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} |g_L(\xi)|^2\,\mathrm{d}\xi = \sum_{n\in\mathbb{Z}} |\hat{f}_L(n)|^2 = \int_{-L/2}^{L/2} |f_L(x)|^2\,\mathrm{d}x.$$

So in the limit as $L\to\infty$ (the period "becomes infinite"), we can think of $g(\xi)$ as some kind of "Fourier transform", since formally:

$$g(\xi) = \lim_{L\to\infty} g_L(\xi) = \int_{-\infty}^{\infty} f(x)e^{-ix\xi}\,\mathrm{d}\xi = \hat{f}(\xi).$$

2.1 Definition and Basic Properties

Definition 2.1. *Let $f \in L^1(\mathbb{R}^n)$. We define $\hat{f}\colon \mathbb{R}^n \to \mathbb{C}$, the* Fourier transform *of f, by*

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x)e^{-2\pi ix\cdot\xi}\,\mathrm{d}x.$$

Proposition 2.2 (Properties of the Fourier transform)**.** *The following properties of the Fourier transform hold:*

(i) (Linearity) *Let $f, g \in L^1(\mathbb{R}^n)$ and $\alpha,\beta\in\mathbb{C}$. Then $(\alpha f + \beta g)\hat{}\,(\xi) = \alpha\hat{f}(\xi) + \beta\hat{g}(\xi)$.*

(ii) (Continuity) *Let $f \in L^1(\mathbb{R}^n)$. Then $\hat{f}$ is continuous, and satisfies $\|\hat{f}\|_{L^\infty} \le \|f\|_{L^1}$.*

(iii) (Riemann-Lebesgue) *Let $f \in L^1(\mathbb{R}^n)$. Then $\lim_{|\xi|\to\infty} |\hat{f}(\xi)| = 0$.*

(iv) (Convolution) *Let $f, g \in L^1(\mathbb{R}^n)$. Then $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$.*

(v) (Shift) *Let $f \in L^1(\mathbb{R}^n)$ and let $h \in \mathbb{R}^n$. For $\tau_h f(x) = f(x+h)$, we have $\widehat{\tau_h f}(\xi) = \hat{f}(\xi)e^{2\pi ih\cdot\xi}$, and for $\sigma_h f(x) = f(x)e^{2\pi ix\cdot h}$, we have $\widehat{\sigma_h f}(\xi) = \hat{f}(\xi - h)$.*

(vi) (Rotation) *Let $f \in L^1(\mathbb{R}^n)$, and let $\Theta \in \mathrm{SO}(n)$ be a rotation matrix. Then $\widehat{f(\Theta\cdot)}(\xi) = \hat{f}(\Theta\xi)$.*

(vii) (Scaling) *Let $f \in L^1(\mathbb{R}^n)$, let $\lambda\in\mathbb{R}$, and define $g(x) = \frac{1}{\lambda^n}f(x/\lambda)$. Then $g \in L^1(\mathbb{R}^n)$, and $\hat{g}(\xi) = \hat{f}(\lambda\xi)$.*

(viii) (Differentiation) *Let* $f \in L^1(\mathbb{R}^n)$ *such that* $\frac{\partial f}{\partial x_j} \in L^1(\mathbb{R}^n)$. *Then* $\widehat{\left(\frac{\partial f}{\partial x_j}\right)}(\xi) = (2\pi i \xi_j)\hat{f}(\xi)$.

(ix) (Multiplication) *Let* $f \in L^1(\mathbb{R}^n)$ *such that* $g_j(x) := -2\pi i x_j f(x)$ *is in* $L^1(\mathbb{R}^n)$. *If* $\hat{f}$ *is differentiable in the* ξ_j *direction, then* $\hat{g}_j(\xi) = \frac{\partial}{\partial \xi_j}\hat{f}(\xi)$.

Proof. We will prove each part individually.

(i) Linearity of the Fourier transform follows from the linearity of the integral.

(ii) For $f \in L^1(\mathbb{R}^n)$, we have that

$$|\hat{f}(\xi)| = \left| \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x \right| \le \int_{\mathbb{R}^n} |f(x)| \, \mathrm{d}x,$$

and hence $\|f\|_{L^\infty} \le \|f\|_{L^1}$. To see that $\hat{f}$ is continuous, for $\xi, h \in \mathbb{R}^n$ consider

$$\hat{f}(\xi + h) - \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) \left(e^{-2\pi i x \cdot (\xi + h)} - e^{-2\pi i x \cdot \xi} \right) \mathrm{d}x.$$

Noticing that the integrand is dominated by $2|f|$, by the dominated convergence theorem the limit as $h \to 0$ exists and equals 0, and hence $\hat{f}$ is continuous.

(iii) This is the analogue to the Riemann–Lebesgue lemma for Fourier series. Suppose that $f \in C^0(\mathbb{R}^n)$. Consider that

$$\begin{aligned} \hat{f}(\xi) &= \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x && (2) \\ &= -\int_{\mathbb{R}^n} f(x) e^{-2\pi i (x + e_n \frac{1}{2\xi_n}) \cdot \xi} \, \mathrm{d}x && \text{where } e_n = (0, \dots, \underbrace{1}_{n^{\text{th}}}, \dots, 0) \\ &= -\int_{\mathbb{R}^n} f(z - \tfrac{1}{2\xi_n} e_n) e^{-2\pi i z \cdot \xi} \, \mathrm{d}z && \text{where } z = x + \frac{1}{2\xi_n} e_n. \quad (3) \end{aligned}$$

From (2) and (3) we obtain that

$$\begin{aligned} \hat{f}(\xi) &= \frac{1}{2} \left(\int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x - \int_{\mathbb{R}^n} f(x - \tfrac{1}{2\xi_n} e_n) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x \right) \\ &= \frac{1}{2} \int_{\mathbb{R}^n} \left(f(x) - f(x - \tfrac{1}{2\xi_n} e_n) \right) e^{-2\pi i x \cdot \xi} \, \mathrm{d}x. \end{aligned}$$

As $|\xi| \to \infty$, we have that $\left(f(x) - f(x - \frac{1}{2\xi_n} e_n) \right) \to 0$ for each $x \in \mathbb{R}^n$ (since f is continuous); as the integrand is dominated by $2|f|$, by the dominated convergence theorem we have that $\lim_{|\xi| \to \infty} |\hat{f}(\xi)| = 0$. This proves the result for all $f \in C^0(\mathbb{R}^n)$.

In general, let $f \in L^1(\mathbb{R}^n)$, and fix $\varepsilon > 0$. As $C^0(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$, pick $g \in C^0(\mathbb{R}^n)$ such that $\|f - g\|_{L^1} < \varepsilon/2$. Then by property (ii), we have that

$\|\hat{f}-\hat{g}\|_{L^\infty} \le \|f-g\|_{L^1} < \varepsilon/2$. Furthermore, as g is continuous, we know that there exists C such that when $|\xi| > C$ we have $|\hat{g}(\xi)| < \varepsilon/2$. Then for $|\xi| > C$, we have

$$\begin{aligned}
|\hat{f}(\xi)| &\le |\hat{f}(\xi) - \hat{g}(\xi)| + |\hat{g}(\xi)| \\
&\le \|\hat{f} - \hat{g}\|_{L^\infty} + |\hat{g}(\xi)| \\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

(iv) The result for convolutions is just an application of Fubini's theorem:

$$\begin{aligned}
\widehat{f*g}(\xi) &= \int_{\mathbb{R}^n} (f*g)(x) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y\, e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x-y)g(y)e^{-2\pi i y\cdot\xi}\,\mathrm{d}y\, e^{2\pi i y\cdot\xi} e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} g(y)e^{-2\pi i y\cdot\xi}\,\mathrm{d}y \int_{\mathbb{R}^n} f(x-y)e^{-2\pi i (x-y)\cdot\xi}\,\mathrm{d}x.
\end{aligned}$$

The result follows after changing variables.

(v) For $\tau_h f(x) = f(x+h)$, we compute that

$$\begin{aligned}
\widehat{\tau_h f}(\xi) &= \int_{\mathbb{R}^n} \tau_h f(x) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x+h) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(z) e^{-2\pi i (z-h)\cdot\xi}\,\mathrm{d}z \\
&= e^{2\pi i h\cdot\xi}\hat{f}(\xi).
\end{aligned}$$

For $\sigma_h f(x) = f(x)e^{2\pi i x\cdot h}$, we compute that

$$\begin{aligned}
\widehat{\sigma_h f}(\xi) &= \int_{\mathbb{R}^n} \sigma_h f(x) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x) e^{2\pi i x\cdot h} e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot(\xi-h)}\,\mathrm{d}x \\
&= \hat{f}(\xi-h).
\end{aligned}$$

(vi) Let $\Theta \in \mathrm{SO}(n)$ be a rotation matrix. As Lebesgue measure is rotationally invariant, we see that

$$\widehat{f(\Theta\cdot)}(\xi) = \int_{\mathbb{R}^n} f(\Theta x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x = \int_{\mathbb{R}^n} f(z)e^{-2\pi i(\Theta^{-1}z)\cdot\xi}\,\mathrm{d}z.$$

Noting that $\Theta^{-1}z\cdot\xi = \Theta^{\mathrm{T}}z\cdot\xi = x\cdot\Theta\xi$, we see that $\widehat{f(\Theta\cdot)}(\xi) = \hat{f}(\Theta\xi)$, as required.

As a corollary of (vi), note that the Fourier transform of a radial function is radial (recall that f is radial if $f(\Theta x) = f(x)$ for all $\Theta \in \mathrm{SO}(n)$), since

$$\hat{f}(\Theta\xi) = \widehat{f(\Theta\cdot)}(\xi) = \hat{f}(\xi).$$

(vii) For $g(x) = \frac{1}{\lambda^n}f(x/\lambda)$, we have that

$$\|g\|_{L^1} = \int_{\mathbb{R}^n} |g(x)|\,\mathrm{d}x = \int_{\mathbb{R}^n} \frac{1}{\lambda^n}|f(x/\lambda)|\,\mathrm{d}x = \int_{\mathbb{R}^n} |f(z)|\,\mathrm{d}z = \|f\|_{L^1},$$

and

$$\begin{aligned}
\hat{g}(\xi) &= \int_{\mathbb{R}^n} g(x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} \frac{1}{\lambda^n} f(x/\lambda)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x/\lambda)e^{-2\pi i (x/\lambda)\cdot\lambda\xi}\,\frac{\mathrm{d}x}{\lambda^n} \\
&= \int_{\mathbb{R}^n} f(y)e^{-2\pi i (y)\cdot\lambda\xi}\,\mathrm{d}y \qquad\qquad \text{putting } y = x/\lambda \\
&= \hat{f}(\lambda\xi).
\end{aligned}$$

(viii) Using integration by parts, we see that

$$\begin{aligned}
\widehat{\left(\frac{\partial f}{\partial x_j}\right)}(\xi) &= \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_j}(x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= 2\pi i\xi_j \int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= 2\pi i\xi_j\hat{f}(\xi).
\end{aligned}$$

(ix) For $g_j(x) := -2\pi i x_j f(x)$, we see that

$$\begin{aligned}
\hat{g}_j(x) &= \int_{\mathbb{R}^n} f(x)(-2\pi i x_j e^{-2\pi i x\cdot\xi})\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} f(x)\frac{\partial}{\partial\xi_j}e^{-2\pi i x\cdot\xi}\,\mathrm{d}x \\
&= \frac{\partial}{\partial\xi_j}\hat{f}(\xi),
\end{aligned}$$

as required. □

From part (ii) of proposition 2.2, we have that $\|\hat{f}\|_{L^\infty} \le \|f\|_{L^1}$. In a finite measure space X (such as on a compact interval such as $[-\pi,\pi]$, or in $\mathbb{T}$), $L^\infty(X) \subsetneq L^1(X)$. However, this is not true in general: $g \in L^\infty(\mathbb{R}^n)$ does *not* imply $g \in L^1(\mathbb{R}^n)$. We would like to be able to say

$$\text{“}f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi)e^{-2\pi i x\cdot\xi}\,\mathrm{d}\xi\text{”},$$

but this makes no sense if f is only in L^1, since we only know that $\hat{f}$ is in L^∞, not L^1. We are thus forced to develop a slightly different theory of the Fourier transform on L^2.

2.2 Schwartz Space and the Fourier Transform

We now consider the class $\mathcal{S}$ of Schwartz functions which are so nice that the Fourier transform of a Schwartz function is another Schwartz function. We have:

$$C_c^\infty(\mathbb{R}^n) \subsetneq \mathcal{S}(\mathbb{R}^n) \subsetneq C^\infty(\mathbb{R}^n).$$

The functions C_c^∞ of compact support are integrable, but there aren't very "many" of them. (Recall that the *support* of a function $f\colon X \to \mathbb{R}$ is defined as $\operatorname{spt} f := \overline{\{x : f(x) \neq 0\}}$, where the line denotes closure, and $f \in C_c^\infty(X)$ if $\operatorname{spt} f$ is compact.) However, move to the larger class of C^∞ functions and you know nothing about integrability. We define a set "between" these two, which is rich enough to contain lots of useful functions, but small enough that we can control the integrability of these functions.

Let us fix some notation: let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and set $|x| = (x_1^2 + \cdots + x_n^2)^{1/2}$. We define a *multi-index* to be an element $\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n$, and write $|\alpha| = \alpha_1 + \cdots + \alpha_n$ and $\alpha! = \alpha_1! \cdot \cdots \cdot \alpha_n!$. We define $x^\alpha := (x_1^{\alpha_1}, \ldots, x_n^{\alpha_n})$, that is, the result of raising each element of x to the corresponding power of α. Observe that $|x^\alpha| \leq c_{n,\alpha}|x|^{|\alpha|}$ for some constant $c_{n,\alpha}$, since for $|x| = 1$ the function $x \mapsto |x^\alpha|$ is continuous on S^{n-1} and hence attains its maximum and minimum, and the result follows by homogeneity of the Euclidean norm. Similarly, for $k \in \mathbb{N}$, we have $|x|^k \leq c'_{n,k} \sum_{|\beta|=k} |x^\beta|$.

If $f\colon \mathbb{R}^n \to \mathbb{C}$ is sufficiently differentiable, we write

$$\partial^\alpha f = \frac{\partial^\alpha f}{\partial x^\alpha} := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}.$$

With this notation, the Leibniz rule for the derivative of a product is

$$\frac{\partial^\alpha (fg)}{\partial x^\alpha} = \sum_{\beta \leq \alpha} \frac{\alpha!}{\beta!(\alpha - \beta)!} \frac{\partial^\beta f}{\partial x^\beta} \frac{\partial^{\alpha-\beta} g}{\partial x^{\alpha-\beta}},$$

where $\beta \leq \alpha$ if and only if $\beta_j \leq \alpha_j$ for each $j = 1, \ldots, n$.

Definition 2.3. *Let $f\colon \mathbb{R}^n \to \mathbb{C}$ be a function in $C^\infty(\mathbb{R}^n)$. For multi-indices α, β, define*

$$\rho_{\alpha,\beta}(f) := \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)|.$$

We define the Schwartz class $\mathcal{S}$ *of functions* $\mathbb{R}^n \to \mathbb{C}$ *as*

$$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) : \rho_{\alpha,\beta}(f) < \infty \text{ for all } \alpha, \beta \in (\mathbb{N} \cup \{0\})^n\}.$$

The $\rho_{\alpha,\beta}$ are seminorms; that is, for all $\alpha, \beta \in (\mathbb{N} \cup \{0\})^n$, $f, g \in \mathcal{S}$, $\lambda, \mu \in \mathbb{C}$, we have

(i) $\rho_{\alpha,\beta}(f) \geq 0$;

(ii) $\rho_{\alpha,\beta}(\lambda f) = |\lambda| \rho_{\alpha,\beta}(f)$, and

(iii) $\rho_{\alpha,\beta}(\lambda f + \mu g) \leq |\lambda| \rho_{\alpha,\beta}(f) + |\mu| \rho_{\alpha,\beta}(g)$.

That is, they are norms *except* that $\rho_{\alpha,\beta}(f) = 0$ does *not* (necessarily) imply that $f = 0$.

It is clear from the definition that, given $f \in \mathcal{S}(\mathbb{R}^n)$, we have that $p(\cdot)f(\cdot) \in \mathcal{S}$ for any polynomial p on $\mathbb{R}^n$, and $\partial^\alpha f \in \mathcal{S}(\mathbb{R}^n)$ for any multi-index α. Note further that $f \in \mathcal{S}(\mathbb{R}^n)$ if, and only if, for every natural number N and every multi-index α, there exists a constant $c_{\alpha,N}$ such that

$$|\partial^\alpha f| \leq c_{\alpha,N} \frac{1}{(1+|x|)^N}. \tag{4}$$

For example, consider $f \in C_c^\infty(\mathbb{R}^n)$. By definition, there exists a compact set K such that $\overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}} \subset K$. Hence each $\rho_{\alpha,\beta}(f)$ is zero outside K, and since every continuous function on a compact set is bounded $\rho_{\alpha,\beta}(f) < +\infty$ and hence that every function in $C_c^\infty(\mathbb{R}^n)$ is in $\mathcal{S}(\mathbb{R}^n)$. As $C_c^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$, we see that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ (for $1 \leq p < \infty$).

However, these are not all the functions in $\mathcal{S}(\mathbb{R}^n)$. For example, the function $x \mapsto e^{-|x|^2}$ is a Schwartz function, since it decays at infinity faster than any polynomial. However, the function $x \mapsto \frac{1}{(1+x^2)^\alpha}$ is *not* in Schwartz space, since multiplying by $x^{3\alpha}$ yields an unbounded function.

Definition 2.4. *Let f_k be a sequence in $\mathcal{S}(\mathbb{R}^n)$, and let $f \in \mathcal{S}(\mathbb{R}^n)$. We say that $f_k \to f$ in $\mathcal{S}(\mathbb{R}^n)$ if, and only if, for every multi-index α and β we have*

$$\rho_{\alpha,\beta}(f_k - f) = \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta (f_k - f)| \to 0$$

as $k \to \infty$.

This defines a topology on $\mathcal{S}(\mathbb{R}^n)$, and addition, scalar multiplication, and differentiation are continuous operators under this topology. What's more, if we let $\{\rho_j\}_{j=1}^\infty$ be some enumeration of the seminorms $\rho_{\alpha,\beta}$, then

$$d(f,g) := \sum_{j=1}^\infty 2^{-j} \frac{\rho_j(f,g)}{1+\rho_j(f,g)}$$

defines a complete metric on $\mathcal{S}(\mathbb{R}^n)$, and $\mathcal{S}(\mathbb{R}^n)$ is locally convex under this metric. Thus $\mathcal{S}(\mathbb{R}^n)$ is an example of a *Fréchet space*: see the appendix to [Fri&Jos] for more details.

Theorem 2.5. *If $f_k \to f$ in $\mathcal{S}(\mathbb{R}^n)$, then for any multi-index β we have that $\partial^\beta f_k \to \partial^\beta f$ in $L^p(\mathbb{R}^n)$ as $k \to \infty$.*

Proof. Set $g_k = f_k - f$. Then

$$\begin{aligned}
\|\partial^\beta g_k\|_{L^p}^p &= \int_{\mathbb{R}^n} |\partial^\beta g_k(x)|^p \,\mathrm{d}x \\
&= \int_{|x|<1} |\partial^\beta g_k(x)|^p \,\mathrm{d}x + \int_{|x|\geq 1} |x|^{n+1} |\partial^\beta g_k(x)|^p \frac{1}{|x|^{n-1}} \,\mathrm{d}x \\
&\leq \|\partial^\beta g_k\|_\infty^p \int_{|x|<1} \mathrm{d}x + \sup_{|x|\geq 1} \left(|x|^{n+1} |\partial^\beta g_k(x)|^p\right) \int_{|x|\geq 1} \frac{1}{|x|^{n-1}} \,\mathrm{d}x \\
&\leq c_{n,p} \left(\|\partial^\beta g_k\|_\infty + \sup_{|x|\geq 1} \left(|x|^{(n+1)/p} |\partial^\beta g_k(x)|\right) \right)^p \\
&\to 0
\end{aligned}$$

as $k \to \infty$. Thus $\partial^\beta f_k \to \partial^\beta f$ in $L^p(\mathbb{R}^n)$, as required. $\square$

We now prove that the Fourier transform maps Schwartz space to Schwartz space, and that the mapping is continuous and invertible.

Theorem 2.6. *The Fourier transform,* $\hat{\cdot}\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$*, given by*

$$\hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot\xi}\,\mathrm{d}x,$$

is a continous linear operator, such that for all $f, g \in \mathcal{S}(\mathbb{R}^n)$ *we have*

$$\int_{\mathbb{R}^n} f(x)\hat{g}(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \hat{f}(x) g(x)\,\mathrm{d}x$$

and that for all $f \in \mathcal{S}(\mathbb{R}^n)$ *and all* $x \in \mathbb{R}^n$ *we have*

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi.$$

In order to prove it, we prove the following lemma:

Lemma 2.7. *If* $f(x) = e^{-\pi|x|^2}$*, then* $\hat{f}(\xi) = e^{-\pi|\xi|^2}$*; that is,* $\hat{f} = f$*.*

Proof. Since f is radial, it suffices to prove this on the real line. As $f'(x) = -2\pi x e^{-\pi x^2}$, notice that $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = e^{-\pi x^2}$ is the unique solution of

$$\begin{cases} u' + 2\pi x u = 0 \\ u(0) = 1 \end{cases} \tag{5}$$

We show that $\hat{f}$ also solves equation (5). First notice that

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x\xi}\,\mathrm{d}x,$$

so

$$\hat{f}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2}\,\mathrm{d}x = 1.$$

We now compute $\hat{f}'$:

$$\begin{aligned}
\hat{f}'(\xi) &= \frac{\mathrm{d}}{\mathrm{d}\xi} \int_{-\infty}^{\infty} e^{-\pi x^2} e^{-2\pi i x\xi}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial\xi} e^{-\pi x^2} e^{-2\pi i x\xi}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} -2\pi i x e^{-\pi x^2} e^{-2\pi i x\xi}\,\mathrm{d}x \\
&= i\widehat{(f')}(\xi) \\
&= -2\pi\xi\hat{f}(\xi)
\end{aligned}$$

by part (viii) of proposition 2.2. Thus $\hat{f} = f$. $\square$

Proof of theorem 2.6. We divide the proof into three parts.

First, we prove that if $f \in \mathcal{S}(\mathbb{R}^n)$ then $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$; that is, for all multi-indices α and β we have that

$$\sup_{\xi \in \mathbb{R}^n} |\xi^\alpha \partial^\beta \hat{f}(\xi)| < \infty.$$

Using parts (viii) and (ix) of proposition 2.2, we see that

$$\xi^\alpha \partial^\beta \hat{f}(\xi) = \xi^\alpha (-2\pi i)^{|\beta|} \widehat{(x^\beta f)} = (2\pi i)^{-|\alpha|} (-2\pi i)^{|\beta|} \widehat{(\partial^\alpha x^\beta f)}.$$

Thus we see that

$$\begin{aligned}
\|\xi^\alpha \partial^\beta \hat{f}(\xi)\|_{L^\infty} &= c \|\widehat{(\partial^\alpha x^\beta f)}\|_{L^\infty} \\
&\leq c \|\partial^\alpha x^\beta f\|_{L^1} \\
&\leq c \left\| \frac{c_{\alpha,N} x^\beta f}{(1+|x|)^N} \right\|_{L^1} && \text{by (4), for all } N \in \mathbb{N} \\
&\leq c \|f\|_{L^\infty} \left\| \frac{x^\beta}{(1+|x|)^N} \right\|_{L^1},
\end{aligned}$$

and $\left\| \frac{x^\beta}{(1+|x|)^N} \right\|_{L^1}$ is finite whenever $N > \beta + n + 1$. Hence $f \in \mathcal{S}(\mathbb{R}^n)$.

For the second part, we prove that, for all $f, g \in \mathcal{S}(\mathbb{R}^n)$, we have

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

This is an application of Fubini's theorem:

$$\begin{aligned}
\int_{\mathbb{R}^n} f(x) \hat{g}(x) \, \mathrm{d}x &= \int_{\mathbb{R}^n} f(x) \int_{\mathbb{R}^n} g(y) e^{-2\pi i x \cdot y} \, \mathrm{d}y \, \mathrm{d}x \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x) g(y) e^{-2\pi i x \cdot y} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{\mathbb{R}^n} g(y) \int_{\mathbb{R}^n} f(x) e^{-2\pi i x \cdot y} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{\mathbb{R}^n} g(y) \hat{f}(y) \, \mathrm{d}y.
\end{aligned}$$

Finally, we prove that for all $f \in \mathcal{S}(\mathbb{R}^n)$ we have

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi.$$

Fix $f \in \mathcal{S}(\mathbb{R}^n)$. Given $g \in \mathcal{S}(\mathbb{R}^n)$, put $g_\lambda(x) = \frac{1}{\lambda^n} g(x/\lambda)$. By the previous part, we have that

$$\int_{\mathbb{R}^n} \hat{f}(x) g_\lambda(x) \, \mathrm{d}x = \int_{\mathbb{R}^n} f(x) \widehat{g_\lambda}(x) \, \mathrm{d}x.$$

By part (vii) of proposition 2.2, we have that $\widehat{g_\lambda}(x) = \hat{g}(\lambda x)$. So,

$$\int_{\mathbb{R}^n} \hat{f}(x) \frac{1}{\lambda^n} g(\tfrac{x}{\lambda}) \, \mathrm{d}x = \int_{\mathbb{R}^n} f(x) \hat{g}(\lambda x) \, \mathrm{d}x = \frac{1}{\lambda^n} \int_{\mathbb{R}^n} f(\tfrac{x}{\lambda}) \hat{g}(x) \, \mathrm{d}x,$$

where the second equality arises from a change of variables $x \mapsto x/\lambda$. Cancelling the $1/\lambda^n$ from both sides, we obtain

$$\int_{\mathbb{R}^n} \hat{f}(x) g(\tfrac{x}{\lambda}) \, \mathrm{d}x = \int_{\mathbb{R}^n} f(\tfrac{x}{\lambda}) \hat{g}(x) \, \mathrm{d}x.$$

By the Dominated Convergence Theorem, as $\lambda \to \infty$, we see that

$$g(0) \int_{\mathbb{R}^n} \hat{f}(x) \, \mathrm{d}x = f(0) \int_{\mathbb{R}^n} \hat{g}(x) \, \mathrm{d}x.$$

In particular, for $g(x) = e^{-\pi|x|^2}$, by lemma 2.7 we have that

$$f(0) = \int_{\mathbb{R}^n} \hat{f}(x) \, \mathrm{d}x;$$

that is, that

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi$$

for $x = 0$. Now, as in part (v) of proposition 2.2, define $\tau_x f(y) = f(y + x)$. Then

$$f(x) = \tau_x f(0) = \int_{\mathbb{R}^n} \widehat{\tau_x f}(\xi) \, \mathrm{d}\xi = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi,$$

by part (v) of proposition 2.2. This completes the proof of the theorem. $\square$

We can thus make the following definition:

Definition 2.8. *Given $f \in \mathcal{S}(\mathbb{R}^n)$, we define the* inverse Fourier transform $\check{f} \in \mathcal{S}(\mathbb{R}^n)$ *by*

$$\check{f}(x) = \int_{\mathbb{R}^n} f(\xi) e^{-2\pi i x \cdot \xi} \, \mathrm{d}\xi.$$

Observe that $\check{f}(x) = \hat{f}(-x)$. Theorem 2.6 thus has the following corollary:

Corollary 2.9. *For all $f \in \mathcal{S}(\mathbb{R}^n)$, we have that $\check{\hat{f}} = \hat{\check{f}} = f$.*

As a consequence of the definition of the Fourier transform for Schwartz functions, we obtain the following two very important results due to Parseval and Plancherel:

Proposition 2.10 (Parseval)**.** *For all $f, g \in \mathcal{S}(\mathbb{R}^n)$, we have that*

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}^n} f(x) \overline{g(x)} \, \mathrm{d}x = \int_{\mathbb{R}^n} \hat{f}(\xi) \overline{\hat{g}(\xi)} \, \mathrm{d}\xi = \langle \hat{f}, \hat{g} \rangle_{L^2}.$$

Proof. Fix $f, g \in \mathcal{S}(\mathbb{R}^n)$. By the second part of theorem 2.6, we have that

$$\int_{\mathbb{R}^n} f(x) \hat{h}(x) \, \mathrm{d}x = \int_{\mathbb{R}^n} \hat{f}(x) h(x) \, \mathrm{d}x$$

for any $h \in \mathcal{S}(\mathbb{R}^n)$. In particular, put $h(x) = \overline{\hat{g}(x)}$. Then we see that

$$\begin{aligned}
\hat{h}(\xi) &= \int_{\mathbb{R}^n} h(x) e^{-2\pi i x\cdot\xi} \, \mathrm{d}x \\
&= \int_{\mathbb{R}^n} \overline{\hat{g}(x)} e^{-2\pi i x\cdot\xi} \, \mathrm{d}x \\
&= \overline{\int_{\mathbb{R}^n} \hat{g}(x) e^{2\pi i x\cdot\xi} \, \mathrm{d}x} \\
&= \overline{g(\xi)}
\end{aligned}$$

by the third part of theorem 2.6. Hence

$$\int_{\mathbb{R}^n} f(x)\overline{g(x)} \, \mathrm{d}x = \int_{\mathbb{R}^n} \hat{f}(x)\overline{\hat{g}(x)} \, \mathrm{d}x. \qquad \square$$

By taking $g = f$, we obtain the following corollary:

Corollary 2.11 (Plancherel)**.** *For all $f \in \mathcal{S}(\mathbb{R}^n)$, we have that*

$$\|f\|_{L^2}^2 = \int_{\mathbb{R}^n} f(x)\overline{f(x)} \, \mathrm{d}x = \int_{\mathbb{R}^n} \hat{f}(\xi)\overline{\hat{f}(\xi)} \, \mathrm{d}\xi = \|\hat{f}\|_{L^2}^2.$$

2.3 Extending the Fourier Transform to $L^p(\mathbb{R}^n)$

Having defined the Fourier transform as a continuous map $\hat{\cdot}\colon \mathcal{S}(\mathbb{R}^n) \to \mathcal{S}(\mathbb{R}^n)$ and shown that it is invertible there, we now seek to extend it to $L^p(\mathbb{R}^n)$ for suitable values of p. We begin by considering the extension to $L^2(\mathbb{R}^n)$.

Corollary 2.11 tells us that $\hat{\cdot}$ is a bounded linear operator on $\mathcal{S}(\mathbb{R}^n) \subset L^1 \cap L^2(\mathbb{R}^n)$, since $\|f\|_{L^2} = \|\hat{f}\|_{L^2}$. As $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^1 \cap L^2(\mathbb{R}^n)$, we can extend the Fourier transform to a unique operator $\hat{\cdot}\colon L^1 \cap L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. Since we are in a subset of $L^1(\mathbb{R}^n)$, the properties of proposition 2.2 all hold.

As $L^1 \cap L^2(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$, we may extend the Fourier transform $\hat{\cdot}\colon L^1 \cap L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ to a unique operator $\mathcal{F}(f)\colon L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. The immediate natural question is, given $f \in L^2(\mathbb{R}^n)$, whether $\int_{\mathbb{R}^n} f(x)e^{-2\pi i x\cdot\xi} \, \mathrm{d}x$ converges, and whether it equals $\mathcal{F}(f)$.

Lemma 2.12. *If $f_k \in L^1 \cap L^2(\mathbb{R}^n)$, and $f_k \to f$ in $L^2(\mathbb{R}^n)$, then $\hat{f}_k$ is a Cauchy sequence in $L^2(\mathbb{R}^n)$.*

Proof. We have

$$\|\hat{f}_j - \hat{f}_k\|_{L^2} = \|(f_j - f_k)\hat{}\|_{L^2} = \|f_j - f_k\|_{L^2},$$

so whenever (f_k) is Cauchy, so is $(\hat{f}_k)$. $\square$

Given a sequence $f_k \in L^1 \cap L^2(\mathbb{R}^n)$ such that $f_k \to f$ for $f \in L^2(\mathbb{R}^n)$, we define $\mathcal{F}(f)$ as the L^2 limit of $\hat{(f_k)}$, that is $\|\hat{f}_k - \mathcal{F}(f)\|_{L^2} \to 0$. (It is easy to check that this is independent of the sequence (f_k) chosen.) For an example of such a sequence, given

$f \in L^2(\mathbb{R}^n)$, we may define $f_k = f\chi_{B_k}$, where $B_k = \{x \in \mathbb{R}^n : |x| \le k\}$ is the ball of radius k about the origin. Note that

$$\int_{\mathbb{R}^n} |f_k(x)|\,\mathrm{d}x = \int_{\mathbb{R}^n} |f(x)|\chi_{B_n}\,\mathrm{d}x \le \|f\|_{L^2}(\operatorname{vol} B_n)^{1/2},$$

so that $f_k \in L^1 \cap L^2(\mathbb{R}^n)$. Since $|f - f_k|^2 = |f|^2(1 - \chi_{B_k}) \le |f|^2$, and $f \in L^2$, by the Dominated Convergence Theorem we have that $f_k \to f$ in $L^2(\mathbb{R}^n)$.

Now, from measure theory we know that if $f_k \to f$ in $L^p(\mu)$, then $f_k \to f$ in measure (with respect to μ); and if $f_k \to f$ in measure (with respect to μ), there exists a subsequence $f_{k_j} \to f$ which converges pointwise μ-almost everywhere. Thus, there exists a sequence k_j such that $\widehat{f_{k_j}} \to \mathcal{F}(f)$ pointwise almost everywhere, i.e. such that

$$\lim_{j\to\infty} \int_{|x|\le k_j} f(x)e^{-2\pi i x\cdot\xi}\,\mathrm{d}x = \mathcal{F}(f)(\xi)$$

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

For the real line, i.e. $n = 1$, it turns out that *all* subsequences converge pointwise almost everywhere, so in fact the *pointwise* (a.e.) limit

$$\lim_{n\to\infty} \hat{f}_n(\xi) = \lim_{n\to\infty} \int_{|x|\le n} f(x)e^{-2\pi i x\xi}\,\mathrm{d}x$$

exists and equals $\mathcal{F}(f)$. It would be nice if the same is true in dimension 2 and higher; however, the answer is not known and this remains an open problem!

One can use the same procedure to define the inverse Fourier transform on $L^2(\mathbb{R}^n)$: denote by $\mathcal{F}'$ the extension of $\check{\ }$ to L^2. Let f_k be a sequence in $\mathcal{S}(\mathbb{R}^n)$ such that $f_k \to f$ in $L^2(\mathbb{R}^n)$; then for each k we know that $\check{f}_k(x) = \hat{f}_k(-x)$, so that $\mathcal{F}'(f)(x) = \mathcal{F}(f)(-x)$ for every $f \in L^2(\mathbb{R}^n)$, as before. By convention, for $f \in L^2$ we write $\hat{f}$ for $\mathcal{F}(f)$, and $\check{f}$ for $\mathcal{F}'(f)$.

Unfortunately, there is no way of extending the Fourier transform to L^1 in such a way that the operation is invertible. So the next question is whether or not we can extend the Fourier transform to $L^p(\mathbb{R}^n)$ for, say, $1 < p < 2$.

Definition 2.13. *Let $1 < p < 2$, and let $f \in L^p(\mathbb{R}^n)$. For a decomposition $f = f_1 + f_2$ where $f_1 \in L^1(\mathbb{R}^n)$ and $f_2 \in L^2(\mathbb{R}^n)$, we define the Fourier transform of f by*

$$\hat{f} = \hat{f}_1 + \hat{f}_2.$$

For an example of such a decomposition, take $f_1 = f\chi_{B_n}$ and $f_2 = f(1 - \chi_{B_n})$.

Lemma 2.14. *Let $1 < p < 2$, and let $f \in L^p(\mathbb{R}^n)$. The Fourier transform of f, as defined above, is independent of the decomposition chosen; that is, if $f = f_1 + f_2 = g_1 + g_2$ are two decompositions, with $f_1, g_1 \in L^1(\mathbb{R}^n)$ and $f_2, g_2 \in L^2(\mathbb{R}^n)$, then $\hat{f}_1 + \hat{f}_2 = \hat{g}_1 + \hat{g}_2$.*

Proof. If $f = f_1 + f_2 = g_1 + g_2$, with $f_1, g_1 \in L^1(\mathbb{R}^n)$ and $f_2, g_2 \in L^2(\mathbb{R}^n)$, then

$$L_1(\mathbb{R}^n) \ni f_1 - g_1 = g_2 - f_2 \in L^2(\mathbb{R}^n),$$

so that $f_1 - g_1, g_2 - f_2 \in L^1 \cap L^2(\mathbb{R}^n)$. Thus we may take their Fourier transform as functions in L^1, and their Fourier transforms will agree, that is $\hat{f}_1 - \hat{g}_1 = \hat{g}_2 - \hat{f}_2$, and hence $\hat{f}_1 + \hat{f}_2 = \hat{g}_1 + \hat{g}_2$, as required. □

Recall that for $f \in L^1(\mathbb{R}^n)$, we have that $\|\hat{f}\|_{L^\infty} \leq \|f\|_{L^1}$; and that for $f \in L^2(\mathbb{R}^n)$ we have that $\|\hat{f}\|_{L^2} \leq \|f\|_{L^2}$ (in fact we have equality, but we ignore that for the time being). We now prove that, for $1 < p < 2$, if $f \in L^p(\mathbb{R}^n)$, we have that $\|\hat{f}\|_{L^q} \leq \|f\|_{L^p}$, whenever $1/p + 1/q = 1$.

Theorem 2.15. *Fix $1 < p < 2$, take q such that $1/p + 1/q = 1$, and let $f \in L^p(\mathbb{R}^n)$. Then $\|\hat{f}\|_{L^q} \leq \|f\|_{L^p}$.*

The proof is an application of the Riesz–Thorin interpolation theorem:

Theorem 2.16 (Riesz–Thorin interpolation theorem)**.** *Let $1 \leq p_0 \leq p_1 \leq \infty$, and $1 \leq q_0 \leq q_1 \leq \infty$. For $\theta \in (0,1)$, define*

$$\begin{aligned} \frac{1}{p} &= \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \\ \frac{1}{q} &= \frac{1-\theta}{q_0} + \frac{\theta}{q_1}. \end{aligned}$$

Note that $p \in (p_0, p_1)$, and $q \in (q_0, q_1)$. Let $T\colon (L^{p_0} + L^{p_1}) \to (L^{q_0} + L^{q_1})$ be a linear operator such that there exist constants M_0 and M_1 such that

$$\begin{aligned} \|Tf\|_{L^{q_0}} &\leq M_0 \|f\|_{L^{p_0}}, \\ \|Tf\|_{L^{q_1}} &\leq M_1 \|f\|_{L^{p_1}}; \end{aligned}$$

i.e. T is bounded as an operator $L^{p_0} \to L^{q_0}$ and as an operator $L^{p_1} \to L^{q_1}$. Then

$$\|Tf\|_{L^q} \leq M_0^{1-\theta} M_1^{\theta} \|f\|_{L^p};$$

that is, T extends uniquely as a bounded linear operator $L^p \to L^q$.

The proof of the Riesz–Thorin theorem uses Hadamard's three-line lemma: if $F\colon S \to \mathbb{C}$ is bounded and continuous on the strip $S = \{x + iy \in \mathbb{C} : 0 \leq x \leq 1\}$, and F is analytic on the interior of S, and there exist constants M_0 and M_1 such that $|F(iy)| \leq M_0$ and $|F(1+iy)| \leq M_1$ for all $y \in \mathbb{R}$, then $|F(x+iy)| \leq M_0^{1-x} M_1^{x}$ for all $x + iy \in S$.

We now use the Riesz–Thorin theorem to prove that the Fourier transform is bounded from $L^p \to L^q$ whenever $1 < p < 2$ and $1/p + 1/q = 1$:

Proof of theorem 2.15. We have that $\hat{\cdot}\colon L^1 \to L^\infty$ and $\hat{\cdot}\colon L^2 \to L^2$ are both bounded. Set $p_0 = 1$, $q_0 = \infty$, and $p_1 = q_1 = 2$. Given $1 < p < 2$, set $\theta = 2 - 2/p$; then

$$\begin{aligned} \frac{1}{p} &= \frac{1-\theta}{1} + \frac{\theta}{2}, \\ \frac{1}{q} &= 1 - \frac{1}{p} = \frac{\theta}{2} = \frac{1-\theta}{\infty} + \frac{\theta}{2}. \end{aligned}$$

Hence, by the Riesz–Thorin theorem, we have that

$$\|\hat{f}\|_{L^q} \leq \|f\|_{L^p}$$

(taking $M_0 = M_1 = 1$), as required. □

As another application of the Riesz–Thorin theorem, we prove Young's theorem about convolutions:

Theorem 2.17 (Young). *Let $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$. If $1 + 1/r = 1/p + 1/q$, then*

$$\|f * g\|_{L^r} \le \|f\|_{L^p} \|g\|_{L^q}.$$

We first prove a lemma, known as Minkowski's inequality:

Lemma 2.18 (Minkowski). *Let $h \colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, and let $1 \le p \le \infty$. Then*

$$\left\| \int_{\mathbb{R}^n} h(x,y)\,\mathrm{d}y \right\|_{L^p_x(\mathbb{R}^n)} \le \int_{\mathbb{R}^n} \|h(x,y)\|_{L^p_x(\mathbb{R}^n)}\,\mathrm{d}y.$$

Proof. If $p = 1$ or $p = \infty$, the result follows from Fubini's theorem. So fix $1 < p < \infty$, and let p' be such that $1/p + 1/p' = 1$. Recall that the dual of L^p, $(L^p)^* \cong L^{p'}$; to calculate the L^p norm of $u \in L^p$, we have that

$$\|u\|_{L^p} = \sup \left\{ \int_{\mathbb{R}^n} u(x)v(x)\,\mathrm{d}x : v \in L^{p'}, \|v\|_{L^{p'}} = 1 \right\}.$$

Fix $v \in L^{p'}$ such that $\|v\|_{L^{p'}} = 1$. Then

$$\begin{aligned}
\left| \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} h(x,y)\,\mathrm{d}y\, v(x)\,\mathrm{d}x \right| &\le \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |h(x,y)||v(x)|\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |h(x,y)||v(x)|\,\mathrm{d}x\,\mathrm{d}y \\
&\le \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |h(x,y)|^p\,\mathrm{d}x \right)^{1/p} \underbrace{\left(\int_{\mathbb{R}^n} |v(x)|^{p'}\,\mathrm{d}x \right)^{1/p'}}_{=1} \mathrm{d}y \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |h(x,y)|^p\,\mathrm{d}x \right)^{1/p} \mathrm{d}y \\
&= \int_{\mathbb{R}^n} \|h(x,y)\|_{L^p_x(\mathbb{R}^n)}\,\mathrm{d}y,
\end{aligned}$$

using Hölder's inequality in x. So taking the supremum over all $v \in L^{p'}$ such that $\|v\|_{L^{p'}} = 1$ we obtain that

$$\begin{aligned}
\left\| \int_{\mathbb{R}^n} h(x,y)\,\mathrm{d}y \right\|_{L^p_x(\mathbb{R}^n)} &= \sup \left\{ \int_{\mathbb{R}^n} \|h(x,y)\|_{L^p_x(\mathbb{R}^n)} v(x)\,\mathrm{d}x : v \in L^{p'}, \|v\|_{L^{p'}} = 1 \right\} \\
&\le \int_{\mathbb{R}^n} \|h(x,y)\|_{L^p_x(\mathbb{R}^n)}\,\mathrm{d}y,
\end{aligned}$$

as required. □

We now proceed to the proof of Young's inequality.

Proof of theorem 2.17. We begin by proving two easy cases. Given p, we denote by p' the number such that $1/p + 1/p' = 1$. First, suppose that $r = \infty$; then $q = p'$, and

$$\begin{aligned}
|f * g| &\leq \int_{\mathbb{R}^n} |f(x-y)g(y)| \,\mathrm{d}y \\
&\leq \left(\int_{\mathbb{R}^n} |f(x-y)|^p \,\mathrm{d}y \right)^{1/p} \left(\int_{\mathbb{R}^n} |g(y)|^{p'} \,\mathrm{d}y \right)^{1/p'} \\
&= \left(\int_{\mathbb{R}^n} |f(z)|^p \,\mathrm{d}z \right)^{1/p} \left(\int_{\mathbb{R}^n} |g(y)|^{p'} \,\mathrm{d}y \right)^{1/p'} \qquad \text{putting } z = x - y \\
&= \|f\|_{L^p} \|g\|_{L^{p'}},
\end{aligned}$$

by Hölder's inequality.

Now, suppose that $r = p$, and hence that $q = 1$. We want to prove that $\|f * g\|_{L^p} \leq \|f\|_{L^p} \|g\|_{L^1}$. Using Minkowski's inequality with $h(x,y) = f(x-y)g(y)$, we have that

$$\begin{aligned}
\|f * g\|_{L^p} &= \left(\int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} f(x-y)g(y) \,\mathrm{d}y \right|^p \mathrm{d}x \right)^{1/p} \\
&\leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f(x-y)g(y)|^p \,\mathrm{d}x \right)^{1/p} \mathrm{d}y \\
&= \int_{\mathbb{R}^n} |g(y)| \left(\int_{\mathbb{R}^n} |f(x-y)|^p \,\mathrm{d}x \right)^{1/p} \mathrm{d}y \\
&= \int_{\mathbb{R}^n} |g(y)| \left(\int_{\mathbb{R}^n} |f(z)|^p \,\mathrm{d}z \right)^{1/p} \mathrm{d}y \qquad \text{putting } z = x - y \\
&= \|f\|_{L^p} \|g\|_{L^1}.
\end{aligned}$$

So, for fixed p, we have proved that Young's inequality holds when $r = \infty$ and $q = p'$, and when $r = p$ and $q = 1$. Fix $f \in L^p(\mathbb{R}^n)$ and define T_f by $T_f(g) = f * g$; we have shown that $T_f \colon L^{p'} \to L^\infty$ is bounded and that $\|T_f g\|_{L^\infty} \leq \|f\|_{L^p} \|g\|_{L^{p'}}$, and that $T_f \colon L^1 \to L^p$ is bounded with $\|T_f g\|_{L^p} \leq \|f\|_{L^p} \|g\|_{L^1}$.

Let (q, r) be a pair satisfying

$$\begin{aligned}
\frac{1}{q} &= \frac{1-\theta}{p'} + \frac{\theta}{1}, \\
\frac{1}{r} &= \frac{1-\theta}{\infty} + \frac{\theta}{p}.
\end{aligned}$$

As $1/p + 1/p' = 1$, we see that $p' = p/(p-1)$. From the second equation we see that $\theta = p/r$, so that

$$\frac{1}{q} = \frac{1 - p/r}{p/(p-1)} + \frac{p}{r} = 1 + \frac{1}{r} - \frac{1}{p}.$$

By the Riesz–Thorin theorem, T_f is bounded as an operator $L^q \to L^r$, so that

$$\|f * g\|_{L^r} \leq \|f\|_{L^p} \|g\|_{L^q}. \qquad \square$$

We have defined the Fourier transform on $L^p(\mathbb{R}^n)$ for $1 \le p \le 2$, but as yet we do not have a simple means for calculating $\hat{f}$, or indeed for reconstructing f from $\hat{f}$. Given $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$, we know that

$$f(x) = \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi.$$

Even if $\hat{f} \notin \mathcal{S}(\mathbb{R}^n)$, but $\hat{f} \in L^2(\mathbb{R}^n)$, then

$$S_R f(x) := \int_{|\xi| \le R} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi$$

always makes sense, since

$$\begin{aligned} \left| \int_{|\xi| \le R} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \, \mathrm{d}\xi \right| &\le \int_{|\xi| \le R} |\hat{f}(\xi)| \, \mathrm{d}\xi \\ &\le (\operatorname{vol} B_r)^{1/2} \left(\int_{|\xi| \le R} |\hat{f}(\xi)|^2 \right)^{1/2} \\ &\le c_n R^{n/2} \|\hat{f}\|_{L^2(\mathbb{R}^n)}. \end{aligned}$$

By analogy with Fourier series, the natural questions to ask are:

1. Does $S_R f \to f$ in the L^p norm?
2. Does $S_R f \to f$ almost everywhere?

The following proposition, which is the analogy of proposition 1.14, is

Proposition 2.19. *Let $1 < p < \infty$. The following are equivalent:*

- *for all $f \in L^p(\mathbb{R}^n)$, $S_R f \to f$ in the L^p norm as $R \to \infty$;*
- *there exists a constant c_p such that, for all $f \in L^p(\mathbb{R}^n)$ and all $R > 0$,*

$$\|S_R f\|_{L^p} \le c_p \|f\|_{L^p}.$$

Proof. First, we show that, if $S_R f \to f$ in the L^p norm for all $f \in L^p(\mathbb{R}^n)$, then such a constant c_p exists. Consider $S_R \colon L^p(\mathbb{R}^n) \to \mathbb{C}$ as operators for each $R > 0$. Each S_R is a bounded linear operator, so by the uniform boundedness principle, either

(i) $\sup_{R>0} \|S_R\|_{\mathrm{op}} < \infty$, or

(ii) there exists $f \in L^p$ such that $\limsup_{R \to \infty} \|S_R f\|_{L^p} = +\infty$.

We show that (i) holds. Given $\varepsilon > 0$ and $f \in L^p(\mathbb{R}^n)$, pick R large enough so that $\|S_R f - f\|_{L^p} < \varepsilon$. Then for such large enough R,

$$\|S_R f\|_{L^p} \le \|S_R f - f\|_{L^p} + \|f\|_{L^p} \le \varepsilon + \|f\|_{L^p}.$$

This holds for all $\varepsilon > 0$, and neither $\|S_R f\|_{L^p}$ nor $\|f\|_{L^p}$ depend on ε, so we have $\|S_R f\|_{L^p} \le \|f\|_{L^p}$ for all f and all R large enough. Hence

$$c_p := \sup_{R>0} \|S_R\|_{\mathrm{op}}$$

exists and is finite.

For the converse — that is, showing that the existence of such a constant c_p guarantees that $S_R f \to f$ in the L^p norm for all $f \in L^p$ — we use the fact that $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ (as long as $1 \leq p < \infty$). Fix $f \in L^p(\mathbb{R}^n)$, and pick $g \in \mathcal{S}(\mathbb{R}^n)$ such that $\|g - f\|_{L^p} < \frac{\varepsilon}{2(1+c_p)}$. We know that $S_R g \to g$ in $L^p(\mathbb{R}^n)$, because $g \in \mathcal{S}(\mathbb{R}^n)$; so pick R big enough such that $\|S_R g - g\|_{L^p} < \varepsilon/2$. Then

$$\begin{aligned}
\|S_R f - f\|_{L^p} &\leq \|S_R f - S_R g\|_{L^p} + \|S_R g - g\|_{L^p} + \|g - f\|_{L^p} \\
&= \|S_R(f - g)\|_{L^p} + \|S_R g - g\|_{L^p} + \|f - g\|_{L^p} \\
&\leq (1 + c_p)\|f - g\|_{L^p} + \|S_R g - g\|_{L^p} \\
&< (1 + c_p)\frac{\varepsilon}{2(1 + c_p)} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

Hence $S_R f \to f$ in the L^p norm, for any $f \in L^p$. □

This reduces convergence in L^p to boundedness in L^p, which is somewhat easier to deal with. It turns out that, in dimension 1, whenever $1 < p \leq 2$ there exists c_p such that

$$\|S_R f\|_{L^p} \leq c_p \|f\|_{L^p}$$

for all $f \in L^p(\mathbb{R})$ and all $R > 0$. Hence $S_R f \to f$ in $L^p(\mathbb{R})$ for $1 < p \leq 2$. However, a famous result of Charles Fefferman from 1971 (which won him the Fields Medal) shows that, in dimension 2 and higher, the inequality

$$\|S_R f\|_{L^2} \leq c_2 \|f\|_{L^2}$$

holds *only* for $p = 2$, and is *false* for all other values of p!

In dimension 1, we can calculate $S_R f$ explicitly:

$$\begin{aligned}
S_R f(x) &= \int_{-R}^{R} \hat{f}(\xi) e^{2\pi i x \xi} \,\mathrm{d}\xi \\
&= \int_{-R}^{R} \int_{-\infty}^{\infty} f(y) e^{-2\pi i y \xi} \,\mathrm{d}y \, e^{2\pi i x \xi} \,\mathrm{d}\xi \\
&= \int_{-\infty}^{\infty} f(y) \int_{-R}^{R} e^{-2\pi i (x-y)\xi} \,\mathrm{d}\xi \,\mathrm{d}y.
\end{aligned}$$

Define the Dirichlet kernel $D_R(t) := \int_{-R}^{R} e^{2\pi i t \xi} \,\mathrm{d}\xi$; then

$$S_R f(x) = \int_{-\infty}^{\infty} f(y) D_R(x - y) \,\mathrm{d}y.$$

We may compute explicitly that

$$\begin{aligned}
D_R(t) &= \int_{-R}^{R} e^{2\pi i t \xi} \,\mathrm{d}\xi \\
&= \frac{1}{2\pi i t}\left(e^{2\pi i R t} - e^{-2\pi i R t}\right) \\
&= \frac{\sin(2\pi R t)}{\pi t}.
\end{aligned}$$

We would like the existence of a constant such that $\|S_R f\|_{L^p} \leq c_p \|f\|_{L^p}$; i.e. such that $\|D_R * f\|_{L^p} \leq c_p \|f\|_{L^p}$. By Young's inequality applied to $D_R * f$, we would like that

$$\|D_R * f\|_{L^p} \leq \|D_R\|_{L^1} \|f\|_{L^p};$$

but unfortunately $\|D_R\|_{L^1} = \infty$. This is somewhat depressing: the "optimal" (in some sense) inequality in L^p spaces doesn't yield the result! However, it will turn out that an improved inequality, namely

$$\|f * g\|_{L^p} \leq \|f\|_{L^p} \|g\|_{L^1_{\mathrm{w}}},$$

where L^1_{w} is a "weak" L^1 space, and fortunately $\|D_R\|_{L^1_{\mathrm{w}}} < \infty$.

2.4 Kernels and PDEs

In a similar vein to Fourier series, we may define various alternative modes of convergence of Fourier transforms, as follows.

Cesàro convergence: Given

$$S_R f(x) := \int_{|\xi| \leq R} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \,\mathrm{d}\xi,$$

form the *Cesàro integral*

$$\sigma_R f(x) = \frac{1}{R} \int_0^R S_r f(x) \,\mathrm{d}r = \int_{\mathbb{R}^n} f(y) \frac{1}{R} \int_0^R D_r(x - y) \,\mathrm{d}r \,\mathrm{d}y.$$

Setting

$$F_R(t) = \frac{1}{R} \int_0^R D_r(t) \,\mathrm{d}r,$$

we can say that $\sigma_R f = F_R * f$. In dimension 1 we have

$$\begin{aligned} F_R(t) &= \frac{1}{R} \int_0^R \frac{\sin(2\pi r t)}{\pi t} \,\mathrm{d}r \\ &= \frac{1 - \cos(2\pi R t)}{2R(\pi t)^2} \\ &= \frac{\sin^2(\pi R t)}{R(\pi t)^2}. \end{aligned}$$

The difference here, however, is that while the norm of the Dirichlet kernel for Fourier series is $\frac{4}{\pi^2} \log N + O(1)$, here we have that $\|D_R\|_{L^1} = +\infty$. However, $\|F_R\|_{L^1} = 1$ for every $R > 0$. So, does $\sigma_R f \to f$ as $R \to \infty$, either in L^p or just almost everywhere?

Abel–Poisson convergence: Given f, we define the *Abel–Poisson integral*

$$u(x, t) = \int_{\mathbb{R}^n} e^{-2\pi t |\xi|} \hat{f}(\xi) e^{2\pi i x \cdot \xi} \,\mathrm{d}\xi.$$

Does $u(x, t) \to f(x)$ as $t \to 0$, either in L^p or almost everywhere?

Gauß–Weierstraß convergence: Given f, we define the *Gauß–Weierstraß integral*

$$\omega(x,t) = \int_{\mathbb{R}^n} e^{-4\pi^2 t|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, \mathrm{d}\xi.$$

Does $\omega(x,t) \to f(x)$ as $t \to 0$, either in L^p or almost everywhere?

The answer in all three of the above cases is yes, to both L^p convergence and almost everywhere convergence! What's more, there are some interesting connections with PDE theory: the Gauß–Weierstraß integral is a solution of the heat equation, and the Abel–Poisson integral is a solution of the Laplace equation.

Proposition 2.20. *Given $f \in L^1(\mathbb{R}^n)$, the Gauß–Weierstraß integral*

$$\omega(x,t) = \int_{\mathbb{R}^n} e^{-4\pi^2 t|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\, \mathrm{d}\xi$$

solves the heat equation on $\mathbb{R}^n \times (0,\infty)$:

$$\begin{cases} \partial_t \omega - \Delta_x \omega = 0 \\ \omega(x,0) = f(x) \end{cases}$$

Heuristic proof. Essentially we take the Fourier transform of the PDE in the x variables, solve the resulting equation, and take the inverse transform to recover the Gauß–Weierstraß integral. First, note that

$$\begin{aligned} \widehat{\partial_t \omega}(\xi,t) &= \int_{\mathbb{R}^n} \partial_t \omega(x,t) e^{-2\pi i x\cdot\xi}\, \mathrm{d}x \\ &= \partial_t \int_{\mathbb{R}^n} \omega(x,t) e^{-2\pi i x\cdot\xi}\, \mathrm{d}x \\ &= \partial_t \hat{\omega}(\xi,t). \end{aligned}$$

By part (viii) of proposition 2.2, we have that

$$\widehat{\left(\frac{\partial \omega}{\partial x_j}\right)}(\xi,t) = (2\pi i \xi_j)\hat{\omega}(\xi,t),$$

so

$$\widehat{\left(\frac{\partial^2 \omega}{\partial x_j^2}\right)}(\xi,t) = (2\pi i \xi_j)^2\hat{\omega}(\xi,t),$$

and hence

$$\widehat{\Delta\omega}(\xi,t) = \sum_{j=1}^{n} -4\pi^2 \xi_j^2 \hat{\omega}(\xi,t) = -4\pi^2|\xi|^2\hat{\omega}(\xi,t).$$

Thus the heat equation becomes

$$\partial_t \hat{\omega}(\xi,t) = -4\pi^2|\xi|^2\hat{\omega}(\xi).$$

This is an ordinary differential equation in t, with solution

$$\hat{\omega}(\xi,t) = e^{-4\pi^2|\xi|^2 t}\hat{f}(\xi).$$

By part (iv) of proposition 2.2, we have that $\widehat{f * g} = \hat{f}\hat{g}$. So if we can find a function W such that $\widehat{W}(\xi, t) = e^{-4\pi^2|\xi|^2 t}$, then we will have that $\omega(x,t) = W * f(x,t)$. We claim that

$$W(x,t) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t}.$$

W is known as the *Weierstraß kernel*. By lemma 2.7, we have that the inverse Fourier transform of $f(\xi) = e^{-\pi|\xi|^2}$ is $\check{f}(x) = e^{-\pi|x|^2}$. So,

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-\pi|\xi|^2(4\pi t)} e^{2\pi x\cdot\xi}\,\mathrm{d}\xi &= \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-\pi|\eta|^2} e^{2\pi(x/\sqrt{4\pi t})\cdot\eta}\,\mathrm{d}\eta \qquad \text{using } \eta = \sqrt{4\pi t}\xi \\
&= \frac{1}{(4\pi t)^{n/2}} e^{-\pi|(x/\sqrt{4\pi t})|^2} \\
&= \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t}.
\end{aligned}$$

Hence

$$\begin{aligned}
\omega(x,t) = W * f(x,t) &= \frac{1}{(4\pi t)^{n/2}} \int_{\mathbb{R}^n} f(y) e^{-|x-y|^2/4t}\,\mathrm{d}y \\
&= (\widehat{W}\widehat{f}\,)^{\vee}(x,t) = \int_{\mathbb{R}^n} e^{-4\pi^2 t|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi
\end{aligned}$$

solves the heat equation. □

Similarly, we have:

Proposition 2.21. *Given $f \in L^1(\mathbb{R}^n)$, the Abel–Poisson integral*

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi.$$

solves the Laplace equation on $\mathbb{R}^n \times (0,\infty)$:

$$\begin{cases} \partial_t^2 u + \Delta_x u = 0 \\ u(x,0) = f(x) \end{cases}$$

Heuristic proof. The proof follows the same lines as the previous proposition. As before, we have that

$$\widehat{\partial_t^2 u}(\xi,t) = \partial_t^2 \hat{u}(\xi,t),$$

and

$$\widehat{\Delta u}(\xi,t) = \sum_{j=1}^n -4\pi^2 \xi_j^2 \hat{u}(\xi,t) = -4\pi^2|\xi|^2 \hat{u}(\xi,t).$$

So the equation becomes

$$\partial_t^2 \hat{u}(\xi,t) = -4\pi^2|\xi|^2 \hat{u}(\xi,t).$$

This is another ODE, this time of second order in t, and it has the solution

$$\hat{u}(\xi,t) = e^{-2\pi|\xi|t}\hat{f}(\xi) + e^{2\pi|\xi|t}\hat{f}(\xi);$$

however, the second term is unbounded so we disregard it and consider only the solution

$$\hat{u}(\xi,t) = e^{-2\pi|\xi|t}\hat{f}(\xi).$$

We now seek $P(x,t)$ such that $\hat{P}(\xi,t) = e^{-2\pi t|\xi|}$, since then by part (iv) of proposition 2.2 we have $u(x,t) = (P * f)(x,t)$. We claim that

$$P(x,t) := \int_{\mathbb{R}^n} e^{-2\pi|\xi|t} e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{t}{(t^2+|x|^2)^{(n+1)/2}}. \tag{6}$$

P is known as the *Poisson kernel.* It suffices to prove (6) for $t = 1$, since if $t \neq 1$ we have

$$\begin{aligned}
\int_{\mathbb{R}^n} e^{-2\pi|\xi|t} e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi &= \frac{1}{t^n}\int_{\mathbb{R}^n} e^{-2\pi|\eta|} e^{2\pi i (x/t)\cdot\eta}\,\mathrm{d}\eta \\
&= \frac{t}{t^{n+1}} \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{(1+|x/t|^2)^{(n+1)/2}} \\
&= \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{t}{(t^2+|x|^2)^{(n+1)/2}}.
\end{aligned}$$

To prove (6) for $t = 1$, we use the "well-known" principle of subordination:

$$e^{-\beta} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} e^{-\beta^2/4u}\,\mathrm{d}u, \tag{7}$$

which we will prove below in lemma 2.22. Applying this with $\beta = 2\pi|\xi|$, we have

$$\begin{aligned}
P(x,1) &= \int_{\mathbb{R}^n} e^{-2\pi|\xi|} e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi \\
&= \int_{\mathbb{R}^n} \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} e^{-4\pi^2|\xi|^2/4u}\,\mathrm{d}u\, e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi && \text{by (7)} \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} \int_{\mathbb{R}^n} e^{-\pi^2|\xi|^2/u} e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi\,\mathrm{d}u && \text{by Fubini's theorem} \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} \left(\frac{u}{\pi}\right)^{n/2} \int_{\mathbb{R}^n} e^{-\pi|\eta|^2} e^{2\pi i (x\sqrt{u/\pi})\cdot\eta}\,\mathrm{d}\eta\,\mathrm{d}u && \text{putting } \eta = \sqrt{\frac{\pi}{u}}\xi \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} \left(\frac{u}{\pi}\right)^{n/2} e^{-u|x|^2}\,\mathrm{d}u && \text{by lemma 2.7} \\
&= \frac{1}{\pi^{(n+1)/2}} \int_0^\infty u^{(n-1)/2} e^{-u(1+|x|^2)}\,\mathrm{d}u \\
&= \frac{1}{\pi^{(n+1)/2}} \frac{1}{(1+|x|^2)^{(n+1)/2}} \int_0^\infty s^{(n-1)/2} e^{-s}\,\mathrm{d}s && \text{putting } s = u(1+|x|^2) \\
&= \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{(1+|x|^2)^{(n+1)/2}},
\end{aligned}$$

as required. □

It remains to prove (7):

Lemma 2.22. *For $\beta \in \mathbb{R}$, we have*

$$e^{-\beta} = \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^u}{\sqrt{u}} e^{-\beta^2/4u} \, \mathrm{d}u.$$

Proof. First, we claim that

$$e^{-\beta} = \frac{2}{\pi} \int_0^\infty \frac{\cos \beta x}{1+x^2} \, \mathrm{d}x.$$

Consider the right-hand side:

$$\frac{2}{\pi} \int_0^\infty \frac{\cos \beta x}{1+x^2} \, \mathrm{d}x = \frac{1}{\pi} \int_{-\infty}^\infty \frac{\cos \beta x}{1+x^2} \, \mathrm{d}x = \frac{1}{\pi} \int_{-\infty}^\infty \frac{e^{i\beta x}}{1+x^2} \, \mathrm{d}x,$$

as sin is an odd function. Consider the contour C in the complex plane given by $[-R, R]$ followed by a semi-circle of radius R in the upper half-plane. Now, $\frac{1}{1+x^2}$ has poles at $x = \pm i$, so

$$\frac{1}{\pi} \int_C \frac{e^{i\beta x}}{1+x^2} \, \mathrm{d}x = 2i \operatorname*{Res}_{x=i} \left(\frac{e^{i\beta x}}{1+x^2} \right) = 2i \operatorname*{Res}_{x=i} \left(\frac{e^{i\beta x}}{(x+i)} \frac{1}{x-i} \right) = 2i \frac{e^{i^2\beta}}{2i} = e^{-\beta}.$$

Since $\frac{e^{i\beta x}}{1+x^2} \to 0$ as $|x| \to +\infty$, we obtain that

$$\frac{1}{\pi} \int_{-\infty}^\infty \frac{e^{i\beta x}}{1+x^2} \, \mathrm{d}x = e^{-\beta},$$

as claimed above.

Now, observe that

$$\int_0^\infty e^{-(1+x^2)u} \, \mathrm{d}u = \frac{1}{1+x^2} \int_0^\infty e^{-s} \, \mathrm{d}s = \frac{1}{1+x^2},$$

using the substitution $s = (1+x^2)u$. Hence, we have that

$$\begin{aligned}
e^{-\beta} &= \frac{2}{\pi} \int_0^\infty \frac{\cos \beta x}{1+x^2} \, \mathrm{d}x \\
&= \frac{2}{\pi} \int_0^\infty \cos \beta x \int_0^\infty e^{-(1+x^2)u} \, \mathrm{d}u \, \mathrm{d}x \\
&= \frac{2}{\pi} \int_0^\infty e^{-u} \int_0^\infty e^{-ux^2} \cos \beta x \, \mathrm{d}x \, \mathrm{d}u && \text{by Fubini's theorem} \\
&= \frac{1}{\pi} \int_0^\infty e^{-u} \int_{-\infty}^\infty e^{-ux^2} \cos \beta x \, \mathrm{d}x \, \mathrm{d}u && \text{by symmetry} \\
&= \frac{1}{\pi} \int_0^\infty e^{-u} \int_{-\infty}^\infty e^{-ux^2} e^{i\beta x} \, \mathrm{d}x \, \mathrm{d}u && \text{as sin is odd} \\
&= \frac{1}{\pi} \int_0^\infty e^{-u} \sqrt{\frac{\pi}{u}} \int_{-\infty}^\infty e^{-\pi y^2} e^{2\pi i (\beta/\sqrt{4\pi u}) y} \, \mathrm{d}y \, \mathrm{d}u && \text{putting } y = x\sqrt{\frac{u}{\pi}} \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\pi (\beta/\sqrt{4\pi u})^2} \, \mathrm{d}u && \text{by lemma 2.7} \\
&= \frac{1}{\sqrt{\pi}} \int_0^\infty \frac{e^{-u}}{\sqrt{u}} e^{-\beta^2/4u} \, \mathrm{d}u,
\end{aligned}$$

as required. □

2.5 Approximations to the Identity

To get back on track, we are considering questions of convergence of Fourier series in the Cesàro, Abel–Poisson and Gauß–Weierstraß senses. We use a technique known as *approximations to the identity* to prove L^p convergence for all three of these modes of convergence.

Definition 2.23 (Approximations to the identity). *Given $\phi \in L^1(\mathbb{R}^n)$ such that $\int_{\mathbb{R}^n} \phi = 1$, we define the* approximation to the identity

$$\phi_t(x) := \frac{1}{t^n}\phi\left(\tfrac{x}{t}\right).$$

For future reference, note that $\phi_t \to \delta$ in $\mathcal{D}'$.

Observe that for $g \in C_c^\infty(\mathbb{R}^n)$, we have

$$\int_{\mathbb{R}^n} \phi_t(x)g(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \frac{1}{t^n}\phi\left(\tfrac{x}{t}\right)g(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \phi(y)g(ty)\,\mathrm{d}y$$

putting $y = x/t$. Therefore

$$\lim_{t\to 0} \int_{\mathbb{R}^n} \phi_t(x)g(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \phi(y)g(0)\,\mathrm{d}y = g(0)$$

as g is continuous, using the Dominated Convergence Theorem.

Lemma 2.24. *Given $\phi \in L^1(\mathbb{R}^n)$ such that $\int_{\mathbb{R}^n} \phi = 1$, and $g \in C_c^\infty(\mathbb{R}^n)$, we have that*

$$\phi_t * g(x) \to g(x)$$

as $t \to 0$.

Proof. First, notice that

$$\phi_t * g(x) = \int_{\mathbb{R}^n} \frac{1}{t^n}\phi\left(\tfrac{y}{t}\right)g(x-y)\,\mathrm{d}y = \int_{\mathbb{R}^n} \phi(z)g(x-tz)\,\mathrm{d}z.$$

By the Dominated Convergence Theorem,

$$\int_{\mathbb{R}^n} \phi(z)g(x-tz)\,\mathrm{d}z \to \int_{\mathbb{R}^n} \phi(z)g(x)\,\mathrm{d}z = g(x)$$

as $t \to 0$, as required. □

In our case, ϕ_t will be one of the Fejér, Poisson or Weierstrass kernels. We will use the following theorem to prove L^p convergence of Fourier series in each of those three modes:

Theorem 2.25. *Let $\phi \in L^1(\mathbb{R}^n)$ such that $\int_{\mathbb{R}^n} \phi = 1$, and fix $1 \le p < \infty$. For any $f \in L^p(\mathbb{R}^n)$, we have that*

$$\|(\phi_t * f) - f\|_{L^p} \to 0$$

as $t \to 0^+$. Moreover, if $f \in C_c^0(\mathbb{R}^n)$, then $\phi_t * f \to f$ uniformly.

Proof. Fix $\varepsilon > 0$. We consider two cases.

Case 1: $f \in L^p(\mathbb{R}^n)$: First, as in the previous lemma, notice that

$$\begin{aligned}\phi_t * f(x) &= \int_{\mathbb{R}^n} \frac{1}{t^n}\phi\left(\tfrac{y}{t}\right) f(x-y)\,\mathrm{d}y \\ &= \int_{\mathbb{R}^n} \phi(z) f(x-tz)\,\mathrm{d}z,\end{aligned}$$

putting $z = y/t$. Hence

$$\begin{aligned}\phi_t * f(x) - f(x) &= \int_{\mathbb{R}^n} \phi(z) f(x-tz)\,\mathrm{d}z - f(x) \\ &= \int_{\mathbb{R}^n} \phi(z)[f(x-tz) - f(x)]\,\mathrm{d}z,\end{aligned}$$

and so

$$\left(\int_{\mathbb{R}^n} |\phi_t * f(x) - f(x)|^p\,\mathrm{d}x\right)^{1/p} = \left(\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} \phi(z)[f(x-tz)-f(x)]\,\mathrm{d}z\right)^p \mathrm{d}x\right)^{1/p}.$$

By Minkowski's inequality (lemma 2.18), we have that

$$\begin{aligned}\left(\int_{\mathbb{R}^n} |\phi_t * f(x) - f(x)|^p\,\mathrm{d}x\right)^{1/p} &\leq \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^n} |\phi(z)|^p |f(x-tz)-f(x)|^p\,\mathrm{d}x\right)^{1/p}\mathrm{d}z \\ &= \int_{\mathbb{R}^n}|\phi(z)|\left(\int_{\mathbb{R}^n} |f(x-tz)-f(x)|^p\,\mathrm{d}x\right)^{1/p}\mathrm{d}z.\end{aligned}$$

As $\phi \in L^1(\mathbb{R}^n)$, we may choose $M > 0$ such that

$$\int_{|z|>M} |\phi(z)|\,\mathrm{d}z \leq \frac{\varepsilon}{4\|f\|_{L^p}}.$$

Then

$$\int_{|z|>M}|\phi(z)|\left(\int_{\mathbb{R}^n}|f(x-tz)-f(x)|^p\,\mathrm{d}x\right)^{1/p}\mathrm{d}z \leq \int_{|z|>M}|\phi(z)|\cdot 2\|f\|_{L^p}\,\mathrm{d}z \leq \frac{\varepsilon}{2}.$$

Now, noting that $|f(x-h)-f(x)| \leq |f(x-h)| + |f(x)|$, by the Dominated Convergence Theorem there exists δ such that whenever $h < M\delta$ we have that

$$\left(\int_{\mathbb{R}^n}|f(x-h)-f(x)|^p\,\mathrm{d}x\right)^{1/p} \leq \frac{\varepsilon}{2}.$$

Then, for $t < \delta$, we have

$$\int_{|z|\leq M}|\phi(z)|\left(\int_{\mathbb{R}^n}|f(x-tz)-f(x)|^p\,\mathrm{d}x\right)^{1/p}\mathrm{d}z \leq \frac{\varepsilon}{2}\int_{|z|\leq M}|\phi(z)|\,\mathrm{d}z \leq \frac{\varepsilon}{2},$$

as $\int_{\mathbb{R}^n} \phi = 1$. Hence for $t < \delta$, we have

$$\begin{aligned}
\|(\phi_t * f) - f\|_{L^p} &= \left(\int_{\mathbb{R}^n} |\phi_t * f(x) - f(x)|^p \, \mathrm{d}x\right)^{1/p} \\
&\leq \int_{\mathbb{R}^n} |\phi(z)| \left(\int_{\mathbb{R}^n} |f(x - tz) - f(x)|^p \, \mathrm{d}x\right)^{1/p} \mathrm{d}z \\
&\leq \int_{|z| \leq M} |\phi(z)| \left(\int_{\mathbb{R}^n} |f(x - tz) - f(x)|^p \, \mathrm{d}x\right)^{1/p} \mathrm{d}z \\
&\quad + \int_{|z| > M} |\phi(z)| \left(\int_{\mathbb{R}^n} |f(x - tz) - f(x)|^p \, \mathrm{d}x\right)^{1/p} \mathrm{d}z \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

as required.

Case 2: $f \in C_c^0(\mathbb{R}^n)$: Once again,

$$\phi_t * f(x) - f(x) = \int_{\mathbb{R}^n} \phi(z)[f(x - tz) - f(x)] \, \mathrm{d}z,$$

and so

$$\begin{aligned}
\sup_{x \in \mathbb{R}^n} |\phi_t * f(x) - f(x)| &= \sup_{x \in \mathbb{R}^n} \left| \int_{\mathbb{R}^n} \phi(z)[f(x - tz) - f(x)] \, \mathrm{d}z \right| \\
&\leq \int_{\mathbb{R}^n} |\phi(z)| \sup_{x \in \mathbb{R}^n} |f(x - tz) - f(x)| \, \mathrm{d}z.
\end{aligned}$$

As $\phi \in L^1(\mathbb{R}^n)$, we may choose $M > 0$ such that

$$\int_{|z| > M} |\phi(z)| \, \mathrm{d}z \leq \frac{\varepsilon}{4\|f\|_\infty}.$$

Then

$$\int_{|z| > M} |\phi(z)| \sup_{x \in \mathbb{R}^n} |f(x - tz) - f(x)| \, \mathrm{d}z \leq \int_{|z| > M} |\phi(z)| \cdot 2\|f\|_\infty \, \mathrm{d}z \leq \frac{\varepsilon}{2}.$$

As $f \in C_c^0(\mathbb{R}^n)$, f is uniformly continuous, and hence there exists $\delta > 0$ such that whenever $h < M\delta$,

$$\sup_{x \in \mathbb{R}^n} |f(x - h) - f(x)| < \frac{\varepsilon}{2}.$$

Then, for $t < \delta$, we have

$$\int_{|z| \leq M} |\phi(z)| \sup_{x \in \mathbb{R}^n} |f(x - tz) - f(x)| \, \mathrm{d}z \leq \frac{\varepsilon}{2} \int_{|z| \leq M} |\phi(z)| \, \mathrm{d}z \leq \frac{\varepsilon}{2},$$

as $\int_{\mathbb{R}^n} \phi = 1$. Hence for $t < \delta$, we have

$$\begin{aligned}
\|(\phi_t * f) - f\|_\infty &= \sup_{x\in\mathbb{R}^n} |\phi_t * f(x) - f(x)| \\
&\leq \int_{\mathbb{R}^n} |\phi(z)| \sup_{x\in\mathbb{R}^n} |f(x - tz) - f(x)| \,\mathrm{d}z \\
&\leq \int_{|z|\leq M} |\phi(z)| \sup_{x\in\mathbb{R}^n} |f(x - tz) - f(x)| \,\mathrm{d}z \\
&\quad + \int_{|z|>M} |\phi(z)| \sup_{x\in\mathbb{R}^n} |f(x - tz) - f(x)| \,\mathrm{d}z \\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,
\end{aligned}$$

as required. $\square$

Corollary 2.26. *If $\phi \in C_c^\infty(\mathbb{R}^n)$ or $\phi \in \mathcal{S}(\mathbb{R}^n)$, then $\phi_t * f$ is smooth. As a consequence, smooth functions are dense in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$.*

Sketch proof. We have that

$$\phi_t * f(x) = \int_{\mathbb{R}^n} \frac{1}{t^n} \phi\left(\tfrac{x-y}{t}\right) f(y) \,\mathrm{d}y,$$

so, differentiating under the integral sign, we obtain

$$\partial_x^\alpha \phi_t * f(x) = \int_{\mathbb{R}^n} \frac{1}{t^n} \partial_x^\alpha \phi\left(\tfrac{x-y}{t}\right) f(y) \,\mathrm{d}y,$$

and the result follows by the Dominated Convergence Theorem. $\square$

We now apply theorem 2.25 to the three modes of convergence discussed above.

Cesàro convergence: We have $\sigma_R f = F_R * f$, where $F_R(x) = \frac{\sin^2(\pi R x)}{R(\pi x)^2}$ in dimension 1. We now prove that $F_R * f$ converges to f in L^p:

Proposition 2.27. *For any $f \in L^p(\mathbb{R})$, we have that*

$$\|(F_R * f) - f\|_{L^p} \to 0$$

as $R \to +\infty$. Moreover, if $f \in C_c^0(\mathbb{R})$, then $F_R * f \to f$ uniformly.

Proof. We wish to check that F_R is an approximation to the identity: that is, that we can write it as $\frac{1}{t}\psi(x/t)$ for some ψ. Let $t = 1/R$, and define

$$\psi(x) := F_1(x) = \frac{\sin^2 \pi x}{(\pi x)^2}.$$

Then

$$F_R(x) = F_{1/t}(x) = \frac{\sin^2(\pi(x/t))}{(1/t)(\pi x)^2} = \frac{\sin^2(\pi(x/t))}{t\pi^2(x/t)^2} = \frac{1}{t}\psi(x/t) = \psi_t(x),$$

so $\sigma_R f = \psi_t * f$. By theorem 2.25, $\|(\sigma_R f) - f\|_{L^p} \to 0$ as $R \to +\infty$. $\square$

Abel–Poisson convergence: We have

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\cdot\xi} \, \mathrm{d}\xi = P_t * f(x),$$

where

$$P_t(x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{t}{(t^2+|x|^2)^{(n+1)/2}}.$$

We now prove that $P_t * f$ converges to f in L^p:

Proposition 2.28. *For any $f \in L^p(\mathbb{R}^n)$, we have that*

$$\|(P_t * f) - f\|_{L^p} \to 0$$

as $t \to 0^+$. Moreover, if $f \in C_c^0(\mathbb{R}^n)$, then $P_t * f \to f$ uniformly.

Proof. We wish to check that P_t is an approximation to the identity: that is, that we can write it as $\frac{1}{t}\psi(x/t)$ for some ψ. Define

$$\psi(x) := P_1(x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{(1+|x|^2)^{(n+1)/2}}.$$

Then

$$P_t(x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{t}{(t^2+|x|^2)^{(n+1)/2}} = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{t^n(1+|x/t|^2)^{(n+1)/2}} = \frac{1}{t^n}\psi(x/t) = \psi_t(x),$$

so $P_t * f = \psi_t * f$. By theorem 2.25, $\|(P_t * f) - f\|_{L^p} \to 0$ as $t \to 0^+$. □

Gauß–Weierstraß convergence: We have

$$\omega(x,t) = \int_{\mathbb{R}^n} e^{-4\pi^2 t|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi} \, \mathrm{d}\xi = W_t * f(x),$$

where

$$W_t(x) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t}.$$

We now prove that $W_t * f$ converges to f in L^p:

Proposition 2.29. *For any $f \in L^p(\mathbb{R}^n)$, we have that*

$$\|(W_t * f) - f\|_{L^p} \to 0$$

as $t \to 0^+$. Moreover, if $f \in C_c^0(\mathbb{R}^n)$, then $W_t * f \to f$ uniformly.

Proof. We wish to check that W_t is an approximation to the identity; however, we cannot do so directly, so we show that $W_{t^2} = \frac{1}{t}\psi(x/t)$ for some ψ. Define

$$\psi(x) := W_1(x) = \frac{1}{(4\pi)^{n/2}} e^{-|x|^2/4}.$$

Then

$$W_{t^2}(x) = \frac{1}{(4\pi t^2)^{n/2}} e^{-|x|^2/4t^2} = \frac{1}{t^n}\frac{1}{(4\pi)^{n/2}} e^{-|x/t|^2/4} = \frac{1}{t^n}\psi(x/t) = \psi_t(x),$$

so $W_{t^2} * f = \psi_t * f$. By theorem 2.25, $\|(W_{t^2} * f) - f\|_{L^p} \to 0$ as $t \to 0^+$, and hence $\|(W_t * f) - f\|_{L^p} \to 0$ as $t \to 0^+$. □

2.6 Weak L^p Spaces

Having shown that the Fourier transform converges in the Cesàro, Abel–Poisson and Gauß–Weierstraß senses in the L^p norm, we now turn our attention to almost everywhere convergence. We would like to have that

$$S_R f(x) = D_R * f(x) = \int_{|\xi|\leq R} \hat{f}(\xi) e^{2\pi i x\cdot\xi}\,\mathrm{d}\xi \to f(x)$$

as $R \to \infty$, but this is only known in dimension 1 and not in higher dimensions. The problem is that $\|D_R\|_{L^1} = +\infty$, so we cannot apply Young's inequality. The underlying problem is that $x \mapsto 1/x$ is not an L^1 function on $\mathbb{R}$; we get round this by creating a replacement space, namely *weak L^p spaces*.

Definition 2.30. *Let $(X, \mathscr{F}, \mu)$ be a measure space, and let $1 \leq p < \infty$. We define the* weak L^p space *$L^p_{\mathrm{w}}(X)$ by*

$$L^p_{\mathrm{w}}(X) := \left\{ f\colon X \to \mathbb{C} \textit{ meas.} : \exists\, c \geq 0 \textit{ s.t. } \forall\, \lambda > 0,\ \mu(\{x \in X : |f(x)| > \lambda\}) \leq \left(\frac{c}{\lambda}\right)^p \right\}.$$

The weak L^p norm *$\|f\|_{L^p_{\mathrm{w}}}$ of a function $f \in L^p_{\mathrm{w}}$ is given by*

$$\|f\|_{L^p_{\mathrm{w}}} := \inf\left\{ c \geq 0 : \forall\, \lambda > 0,\ \mu(\{x \in X : |f(x)| > \lambda\}) \leq \left(\frac{c}{\lambda}\right)^p \right\}$$

i.e. the least constant c such that the above inequality is satisfied. For $p = \infty$, we define $L^\infty_{\mathrm{w}} := L^\infty$, and $\|f\|_{L^\infty_{\mathrm{w}}} := \|f\|_{L^\infty}$.

Note that L^p_{w} is a vector space: given $f, g \in L^p_{\mathrm{w}}$, we see that

$$\begin{aligned}
\mu(\{x \in X : |f(x) + g(x)| > \lambda\}) &\leq \mu(\{x \in X : |f(x)| > \tfrac{\lambda}{2}\}) + \mu(\{x \in X : |g(x)| > \tfrac{\lambda}{2}\}), \\
&\leq \left(\frac{2\|f\|_{L^p_{\mathrm{w}}}}{\lambda}\right)^p + \left(\frac{2\|g\|_{L^p_{\mathrm{w}}}}{\lambda}\right)^p \\
&= \left(\frac{2(\|f\|^p_{L^p_{\mathrm{w}}} + \|g\|^p_{L^p_{\mathrm{w}}})^{1/p}}{\lambda}\right)^p,
\end{aligned}$$

so that $\|f + g\|_{L^p_{\mathrm{w}}} \leq 2(\|f\|^p_{L^p_{\mathrm{w}}} + \|g\|^p_{L^p_{\mathrm{w}}})^{1/p}$. Furthermore, given $\alpha \in \mathbb{C}$ and $f \in L^p_{\mathrm{w}}$, it is clear that

$$\mu(\{x \in X : |\alpha f(x)| > \lambda\}) = \mu(\{x \in X : |f(x)| > \tfrac{\lambda}{|\alpha|}\}),$$

so that $\|\alpha f\|_{L^p_{\mathrm{w}}} = |\alpha|\|f\|_{L^p_{\mathrm{w}}}$.

It would be slightly ridiculous if we had defined weak L^p spaces in such a way as an L^p function was not necessarily a weak L^p function, so we now verify that $L^p(X) \subset L^p_{\mathrm{w}}(X)$.

Lemma 2.31. *Let $(X, \mathscr{F}, \mu)$ be a measure space, fix $1 \leq p < \infty$, and let $f \in L^p(X)$. Then $f \in L^p_{\mathrm{w}}$, and $\|f\|_{L^p_{\mathrm{w}}} \leq \|f\|_{L^p}$.*

Proof. Let $f \in L^p(X)$, and set $E_\lambda := \{x \in X : |f(x)| > \lambda\}$. Then

$$\|f\|^p_{L^p} = \int_X |f|^p \,\mathrm{d}\mu \geq \int_{E_\lambda} |f|^p \,\mathrm{d}\mu \geq \int_{E_\lambda} \lambda^p \,\mathrm{d}\mu = \lambda^p \mu(E_\lambda),$$

so that

$$\mu(E_\lambda) \leq \left(\frac{\|f\|_{L^p}}{\lambda}\right)^p.$$

Hence $f \in L^p_{\mathrm{w}}$, and $\|f\|_{L^p_{\mathrm{w}}} \leq \|f\|_{L^p}$, as required. $\square$

In general, $L^p(X) \subsetneq L^p_{\mathrm{w}}(X)$; that is, there are functions in L^p_{w} which are not in L^p. For example, consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = 1/x$. Then $f \notin L^1(\mathbb{R})$, but $f \in L^1_{\mathrm{w}}(\mathbb{R})$, as

$$\mu\left(\left\{x : \frac{1}{|x|} > \lambda\right\}\right) = \mu\left(\left\{x : \frac{1}{\lambda} > |x| \geq 0\right\}\right) = \mu\left(\left[-\frac{1}{\lambda}, \frac{1}{\lambda}\right]\right) = \frac{2}{\lambda}.$$

In higher dimensions, one can see that $x \mapsto 1/|x|^n$ is in $L^1_{\mathrm{w}}(\mathbb{R}^n)$, but not $L^1(\mathbb{R}^n)$, and that $x \mapsto 1/|x|^{n/p}$ is in $L^p_{\mathrm{w}}(\mathbb{R}^n)$, but not $L^p(\mathbb{R}^n)$.

Beware: the notation $\|\cdot\|_{L^p_{\mathrm{w}}}$ is a bit misleading, since the weak L^p norm is *not* a norm, because it does not satisfy the triangle inequality, though all the other axioms of a norm are satisfied. Consider $f(x) = \frac{1}{x}$ and $g(x) = \frac{1}{1-x}$ on the real line: then $\|f\|_{L^1_{\mathrm{w}}} = 2$ and $\|g\|_{L^1_{\mathrm{w}}} = 2$ (as above), but $\|f+g\|_{L^1_{\mathrm{w}}} = 4\sqrt{2} > 4$. To see this, first note that $\frac{1}{x} + \frac{1}{1-x} = \frac{1}{x(1-x)}$, and set $E_\lambda := \left\{x : \frac{1}{|x(1-x)|} > \lambda\right\}$. Then

$$\begin{aligned} E_\lambda &= \left\{x : \frac{1}{\lambda} > |x(1-x)| \geq 0\right\} \\ &= \begin{cases} \left(\frac{1}{2} - \frac{\sqrt{1+4/\lambda}}{2}, \frac{1}{2} + \frac{\sqrt{1+4/\lambda}}{2}\right) & \text{if } \lambda < 4 \\ \left(\frac{1}{2} - \frac{\sqrt{1+4/\lambda}}{2}, \frac{1}{2} - \frac{\sqrt{1-4/\lambda}}{2}\right) \cup \left(\frac{1}{2} + \frac{\sqrt{1-4/\lambda}}{2}, \frac{1}{2} + \frac{\sqrt{1+4/\lambda}}{2}\right) & \text{if } \lambda \geq 4 \end{cases} \end{aligned}$$

and hence we see that

$$\mu(E_\lambda) = \begin{cases} \sqrt{1+4/\lambda} & \text{if } \lambda < 4 \\ \sqrt{1+4/\lambda} - \sqrt{1-4/\lambda} & \text{if } \lambda \geq 4 \end{cases}$$

It is then clear that the maximum value of $\lambda\mu(E_\lambda)$ occurs with $\lambda = 4$ (simply note that $\lambda \mapsto \lambda\mu(E_\lambda)$ is increasing for $\lambda < 4$ and decreasing for $\lambda > 4$). Then $\mu(E_4) = \sqrt{2}$, so that the smallest constant must be $4\sqrt{2}$, and hence $\|f+g\|_{L^1_{\mathrm{w}}} = 4\sqrt{2}$.

Having defined weak L^p spaces, we now consider operators between such spaces, and their boundedness. An operator $T\colon L^p(X) \to L^q(X)$ is bounded if

$$\|Tf\|_{L^q} \leq c\|f\|_{L^p}$$

for all $f \in L^p(X)$. We now consider the more general case where an operator may be defined on some superset of $L^p(X)$: let us denote by $\mathcal{L}(X)$ the space of all measurable functions on X. We may still want to talk about boundedness as an operator $L^p(X) \to L^q(X)$, or we may want to talk about boundedness as an operator $L^p(X) \to L^q_{\mathrm{w}}(X)$: these two possibilities are encapsulated by saying an operator is *strong*-(p,q) or *weak*-(p,q).

Definition 2.32. *Let $(X, \mathscr{F}, \mu)$ be a σ-finite measure space, and consider an operator $T\colon \mathcal{L}(X) \to \mathcal{L}(X)$.*

- *T is* strong-(p,q) *if T is bounded as an operator $L^p(X) \to L^q(X)$, that is there is a constant c such that*
$$\|Tf\|_{L^q} \leq c\|f\|_{L^p}$$
for all $f \in L^p(X)$.

- *T is* weak-(p,q) *if T is bounded as an operator $L^p(X) \to L^q_{\mathrm{w}}(X)$, that is there is a constant c such that*
$$\|Tf\|_{L^q_{\mathrm{w}}} \leq c\|f\|_{L^p}$$
for all $f \in L^p(X)$. If $q = \infty$ this is the same as strong-(p,q); if $q < \infty$, this is equivalent to
$$\mu(\{x \in X : |Tf(x)| > \lambda\}) \leq \left(\frac{c\|f\|_{L^p}}{\lambda}\right)^q$$
for all $f \in L^p(X)$ and all $\lambda > 0$.

Lemma 2.33. *If T is strong-(p,q), then T is weak-(p,q).*

(This follows immediately from the previous lemma, but it is instructive to work through the proof in detail.)

Proof. Suppose T is strong-(p,q), let $f \in L^p(X)$, and set $E_\lambda := \{x \in X : |Tf(x)| > \lambda\}$. Then
$$(c\|f\|_{L^p})^q \geq \|Tf\|_{L^q}^q = \int_X |Tf|^q \,\mathrm{d}\mu \geq \int_{E_\lambda} |Tf|^q \,\mathrm{d}\mu \geq \int_{E_\lambda} \lambda^q \,\mathrm{d}\mu = \lambda^q \mu(E_\lambda),$$
so that
$$\mu(E_\lambda) \leq \left(\frac{c\|f\|_{L^p}}{\lambda}\right)^q.$$
Hence T is weak-(p,q), as required. $\square$

Many of the operators we come across will not be strong-(p,q), but rather weak-(p,q). The Marcinkiewicz interpolation theorem tells us that, if we know that T is weak-(p_0,p_0) and weak-(p_1,p_1), then T is strong-(p,p) for all $p_0 < p < p_1$.

Theorem 2.34 (Marcinkiewicz interpolation theorem)**.** *Let $(X, \mathscr{F}, \mu)$ be a σ-finite measure space, and let $T\colon \mathcal{L}(X) \to \mathcal{L}(X)$. Suppose that T is sublinear, i.e. for all $f, g \in \mathcal{L}(X)$, and all $\alpha \in \mathbb{C}$, we have that*
$$|T(f+g)(x)| \leq |Tf(x)| + |Tg(x)|,$$
$$|T(\alpha f)(x)| = |\alpha||Tf(x)|.$$
If there are $1 \leq p_0 < p_1 \leq \infty$ such that T is weak-(p_0,p_0) and weak-(p_1,p_1), then T is strong-(p,p) for all $p_0 < p < p_1$.

The Marcinkiewicz interpolation theorem saves us a lot of work in proving boundedness of an operator: instead of having to prove boundedness for all $1 < p < \infty$, it suffices to prove that the operator is weak-$(1,1)$ and weak-(∞,∞).

Given a measurable function $f\colon X \to \mathbb{C}$, define

$$a_f(\lambda) := \mu\left(\{x \in X : |f(x)| > \lambda\}\right).$$

Proposition 2.35. *Let $\phi\colon [0,\infty) \to [0,\infty)$ be a differentiable increasing function such that $\phi(0) = 0$. Then*

$$\int_X \phi(|f(x)|)\,\mathrm{d}\mu = \int_0^\infty \phi'(\lambda)a_f(\lambda)\,\mathrm{d}\lambda.$$

Proof. Using the fundamental theorem of calculus, we compute that

$$\begin{aligned}
\int_X \phi(|f(x)|)\,\mathrm{d}\mu &= \int_X \int_0^{|f(x)|} \phi'(\lambda)\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= \int_X \int_0^\infty \phi'(\lambda)\chi_{\{0\le\lambda\le|f(x)|\}}(\lambda,x)\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= \int_0^\infty \phi'(\lambda)\underbrace{\int_X \chi_{\{0\le\lambda\le|f(x)|\}}(\lambda,x)\,\mathrm{d}\mu}_{=\mu(\{x\in X:|f(x)|>\lambda\})}\,\mathrm{d}\lambda \\
&= \int_0^\infty \phi'(\lambda)a_f(\lambda)\,\mathrm{d}\lambda.
\end{aligned}$$

□

Applying the proposition to $\phi(x) = x^p$, for $1 \le p < \infty$, noting that $\phi'(x) = px^{p-1}$, we obtain that

$$\|f\|_{L^p}^p = \int_X |f(x)|^p\,\mathrm{d}x = \int_0^\infty p\lambda^{p-1}a_f(\lambda)\,\mathrm{d}\lambda = \int_0^\infty p\lambda^{p-1}\mu(\{x \in X : |f(x)| > \lambda\})\,\mathrm{d}\lambda.$$

We use this observation in the proof of the Marcinkiewicz interpolation theorem:

Proof of theorem 2.34. Given $f \in L^p(X)$, with $p_0 < p < p_1$, for each $\lambda > 0$ we decompose f as

$$\begin{aligned}
f_0 &= f \cdot \chi_{\{x\in X:|f|>c\lambda\}} \in L^{p_0}, \\
f_1 &= f \cdot \chi_{\{x\in X:|f|\le c\lambda\}} \in L^{p_1},
\end{aligned}$$

so that $f = f_0 + f_1$, with the constant c to be chosen below. As T is sublinear, we have that

$$|Tf(x)| \le |Tf_0(x)| + |Tf_1(x)|,$$

and hence that

$$\mu\left(\{x \in X : |Tf(x)| > \lambda\}\right) \le \mu\left(\left\{x \in X : |Tf_0(x)| > \tfrac{\lambda}{2}\right\}\right) + \mu\left(\left\{x \in X : |Tf_1(x)| > \tfrac{\lambda}{2}\right\}\right),$$

or, put more briefly,

$$a_{Tf}(\lambda) \le a_{Tf_0}(\tfrac{\lambda}{2}) + a_{Tf_1}(\tfrac{\lambda}{2}).$$

We want to show that $\|Tf\|_{L^p} \leq c\|f\|_{L^p}$: the remainder of the proof is in two cases, depending on whether $p_1 = \infty$ or $p_1 < \infty$.

Case 1: $p_1 = \infty$. We know that T is weak-(p_0, p_0), so that there exists A_0 such that, for all $f = f_0 + f_1$ and all $\lambda > 0$, we have

$$a_{Tf_0}(\lambda) \leq \left(\frac{A_0\|f_0\|_{L^{p_0}}}{\lambda}\right)^{p_0}.$$

Also, we know that T is weak-(∞, ∞), so there exists A_1 such that, for all $f = f_0 + f_1$ and all $\lambda > 0$, we have

$$\|Tf_1\|_{L^\infty} \leq A_1\|f_1\|_{L^\infty}.$$

If we choose the constant c above to be $c = \frac{1}{2A_1}$, then

$$\begin{aligned}
a_{Tf_1}(\tfrac{\lambda}{2}) &= \mu\left(\left\{x \in X : |Tf_1(x)| > \tfrac{\lambda}{2}\right\}\right) \\
&\leq \mu\left(\left\{x \in X : A_1|f_1(x)| > \tfrac{\lambda}{2}\right\}\right) \\
&\leq \mu\left(\left\{x \in X : A_1 c\lambda > \tfrac{\lambda}{2}\right\}\right) \\
&= 0,
\end{aligned}$$

since $\frac{\lambda}{2} > \frac{\lambda}{2}$ is impossible. Hence, using the remark following proposition 2.35, we obtain that

$$\begin{aligned}
\|Tf\|_{L^p}^p &= \int_0^\infty p\lambda^{p-1} a_{Tf}(\lambda)\,\mathrm{d}\lambda \\
&\leq \int_0^\infty p\lambda^{p-1} a_{Tf_0}(\tfrac{\lambda}{2})\,\mathrm{d}\lambda \\
&\leq \int_0^\infty p\lambda^{p-1}\left(\frac{2A_0\|f_0\|_{L^{p_0}}}{\lambda}\right)^{p_0}\mathrm{d}\lambda \\
&= (2A_0)^{p_0} p \int_0^\infty \lambda^{p-1-p_0} \int_X |f_0(x)|^{p_0}\,\mathrm{d}\mu\,\mathrm{d}\lambda \\
&= (2A_0)^{p_0} p \int_0^\infty \lambda^{p-1-p_0} \int_X |f(x)|^{p_0} \chi_{\{x\in X : |f(x)| > \frac{\lambda}{2A_1}\}}(x)\,\mathrm{d}\mu\,\mathrm{d}\lambda \\
&= (2A_0)^{p_0} p \int_X |f(x)|^{p_0} \int_0^\infty \lambda^{p-1-p_0} \chi_{\{x\in X : |f(x)| > \frac{\lambda}{2A_1}\}}(x)\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} p \int_X |f(x)|^{p_0} \int_0^{2A_1|f(x)|} \lambda^{p-1-p_0}\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} \frac{p}{p-p_0} \int_X |f(x)|^{p_0} (2A_1|f(x)|)^{p-p_0}\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} (2A_1)^{p-p_0} \frac{p}{p-p_0} \int_X |f(x)|^p\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} (2A_1)^{p-p_0} \frac{p}{p-p_0} \|f\|_{L^p}^p,
\end{aligned}$$

as required.

Case 2: $p_1 < \infty$. As T is weak-(p_0, p_0) and weak-(p_1, p_1), there exist constants A_0 and A_1 such that, for all $f = f_0 + f_1$ and all $\lambda > 0$, we have

$$a_{Tf_0}(\lambda) \leq \left(\frac{A_0\|f_0\|_{L^{p_0}}}{\lambda}\right)^{p_0},$$

and

$$a_{Tf_1}(\lambda) \leq \left(\frac{A_1\|f_1\|_{L^{p_1}}}{\lambda}\right)^{p_1}.$$

(Note that the constants A_0 and A_1 are independent of f.) Using the remark following proposition 2.35, we obtain that

$$\begin{aligned}
\|Tf\|_{L^p}^p &= \int_0^\infty p\lambda^{p-1} a_{Tf}(\lambda)\,\mathrm{d}\lambda \\
&\leq \int_0^\infty p\lambda^{p-1} a_{Tf_0}(\tfrac{\lambda}{2})\,\mathrm{d}\lambda + \int_0^\infty p\lambda^{p-1} a_{Tf_1}(\tfrac{\lambda}{2})\,\mathrm{d}\lambda \\
&\leq \int_0^\infty p\lambda^{p-1}\left(\frac{2A_0\|f_0\|_{L^{p_0}}}{\lambda}\right)^{p_0}\mathrm{d}\lambda + \int_0^\infty p\lambda^{p-1}\left(\frac{2A_1\|f_1\|_{L^{p_1}}}{\lambda}\right)^{p_1}\mathrm{d}\lambda \\
&= (2A_0)^{p_0} p \int_0^\infty \lambda^{p-1-p_0} \int_X |f(x)|^{p_0} \chi_{\{x\in X : |f(x)| > c\lambda\}}\,\mathrm{d}\mu\,\mathrm{d}\lambda \\
&\quad + (2A_1)^{p_1} p \int_0^\infty \lambda^{p-1-p_1} \int_X |f(x)|^{p_1} \chi_{\{x\in X : |f(x)| \leq c\lambda\}}\,\mathrm{d}\mu\,\mathrm{d}\lambda \\
&= (2A_0)^{p_0} p \int_X |f(x)|^{p_0} \int_0^\infty \lambda^{p-1-p_0} \chi_{\{x\in X : |f(x)| > c\lambda\}}\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&\quad + (2A_1)^{p_1} p \int_X |f(x)|^{p_1} \int_0^\infty \lambda^{p-1-p_1} \chi_{\{x\in X : |f(x)| \leq c\lambda\}}\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} p \int_X |f(x)|^{p_0} \int_0^{|f(x)|/c} \lambda^{p-1-p_0}\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&\quad + (2A_1)^{p_1} p \int_X |f(x)|^{p_1} \int_{|f(x)|/c}^\infty \lambda^{p-1-p_1}\,\mathrm{d}\lambda\,\mathrm{d}\mu \\
&= (2A_0)^{p_0} \frac{p}{p-p_0} \int_X |f(x)|^{p_0} (|f(x)|/c)^{p-p_0}\,\mathrm{d}\mu \\
&\quad + (2A_1)^{p_1} \frac{p}{p_1-p} \int_X |f(x)|^{p_1} (|f(x)|/c)^{p-p_1}\,\mathrm{d}\mu \\
&= \left(\frac{(2A_0)^{p_0} p}{c^{p-p_0}(p-p_0)} + \frac{(2A_1)^{p_1} p}{c^{p-p_1}(p_1-p)}\right) \int_X |f(x)|^p\,\mathrm{d}\mu \\
&= \left(\frac{(2A_0)^{p_0} p}{c^{p-p_0}(p-p_0)} + \frac{(2A_1)^{p_1} p}{c^{p-p_1}(p_1-p)}\right) \|f\|_{L^p}^p,
\end{aligned}$$

as required. (Notice that when we integrated λ^{p-1-p_0} and λ^{p-1-p_1}, we used the fact that, since $p - p_0 > 0$, $\lambda^{p-p_0} \to 0$ as $\lambda \to 0$, and, since $p - p_1 < 0$, $\lambda^{p-p_1} \to 0$ as $\lambda \to +\infty$.) This completes the proof of the Marcinkiewicz interpolation theorem. $\square$

2.7 Maximal Functions and Almost Everywhere Convergence

Having proved the Marcinkiewicz interpolation theorem, we now turn our attention to how to relate the weak L^p spaces to almost everywhere convergence. This is done through *maximal functions.*

Theorem 2.36. *Let $T_t\colon L^p(X) \to \mathcal{L}(X)$ be a family of operators for $t > 0$. Define $T^*\colon L^p(X) \to \mathcal{L}(X)$ by*

$$T^* f(x) := \sup_{t>0} |T_t f(x)|.$$

If there is q such that T^* is weak-(p, q), then the set

$$\{f \in L^p(X) : \lim_{t\to 0} T_t f(x) = f(x) \text{ for a.e. } x \in X\}$$

is closed in $L^p(X)$.

T^* is known as the *maximal operator* associated to the family T_t.

Proof. Let $f_n, f \in L^p(X)$ such that $\|f_n - f\|_{L^p} \to 0$ as $n \to \infty$, and suppose that, for every $n \in \mathbb{N}$,

$$\lim_{t\to 0} T_t f_n(x) = f_n(x) \quad \text{for a.e. } x \in X.$$

We want to show that $\lim_{t\to 0} T_t f(x) = f(x)$ for almost every $x \in X$.

Fix $\lambda > 0$. Then

$$\begin{aligned}
&\mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x) - f(x)| > \lambda\right\}\right)\\
&\quad\le \mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t(f - f_n)(x) - (f - f_n)(x)| > \lambda\right\}\right)\\
&\quad\le \mu\left(\left\{x \in X : \left[\sup_{t\to 0} |T_t(f - f_n)(x)| + |(f - f_n)(x)|\right] > \lambda\right\}\right)\\
&\quad\le \mu\left(\left\{x \in X : \sup_{t\to 0} |T_t(f - f_n)(x)| > \frac{\lambda}{2}\right\}\right) + \mu\left(\left\{x \in X : |(f - f_n)(x)| > \frac{\lambda}{2}\right\}\right)\\
&\quad= \mu\left(\left\{x \in X : T^*(f - f_n)(x) > \frac{\lambda}{2}\right\}\right) + \mu\left(\left\{x \in X : |(f - f_n)(x)| > \frac{\lambda}{2}\right\}\right)\\
&\quad\le \left(\frac{2c}{\lambda}\|f - f_n\|_{L^p}\right)^q + \left(\frac{2}{\lambda}\|f - f_n\|_{L^p}\right)^p
\end{aligned}$$

as T^* is weak-(p, q), and $\|f - f_n\|_{L^p_{\mathrm{w}}} \le \|f - f_n\|_{L^p}$. The last line tends to 0 as $n \to \infty$, and since the first line is independent of n we must have that

$$\mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x) - f(x)| > \lambda\right\}\right) = 0$$

for all $\lambda > 0$. Hence

$$\begin{aligned}
\mu&\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x) - f(x)| > 0\right\}\right) \\
&\le \mu\left(\bigcup_{k=1}^{\infty}\left\{x \in X : \limsup_{t\to 0} |T_t f(x) - f(x)| > \frac{1}{k}\right\}\right) \\
&\le \sum_{k=1}^{\infty} \mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x) - f(x)| > \frac{1}{k}\right\}\right) \\
&= 0,
\end{aligned}$$

so that $\lim_{t\to 0} T_t f(x) = f(x)$ for almost every $x \in X$. Hence the set

$$\{f \in L^p(X) : \lim_{t\to 0} T_t f(x) = f(x) \text{ for a.e. } x \in X\}$$

is closed in $L^p(X)$. $\square$

By a very similar argument, we obtain the following corollary:

Corollary 2.37. *Let $T_t\colon L^p(X) \to \mathcal{L}(X)$ be a family of operators for $t > 0$. Define $T^*\colon L^p(X) \to \mathcal{L}(X)$ by*

$$T^* f(x) := \sup_{t>0} |T_t f(x)|.$$

If there is q such that T^* is weak-(p,q), then the set

$$\{f \in L^p(X) : \lim_{t\to 0} T_t f(x) \textit{ exists for a.e. } x \in X\}$$

is closed in $L^p(X)$.

Proof. Let $f_n, f \in L^p(X)$ such that $\|f_n - f\|_{L^p} \to 0$ as $n \to \infty$, and suppose that, for every $n \in \mathbb{N}$, $\lim_{t\to 0} T_t f_n(x)$ exists for almost every $x \in X$. As in the proof of the previous theorem, it suffices to show that

$$\mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x)| - \liminf_{t\to 0} |T_t f(x)| > \lambda\right\}\right) = 0$$

for all $\lambda > 0$. To see this, observe that

$$\limsup_{t\to 0} |T_t f(x)| - \liminf_{t\to 0} |T_t f(x)| \le 2T^* f(x).$$

Then

$$\begin{aligned}
\mu&\left(\left\{x \in X : \limsup_{t\to 0} |T_t f(x)| - \liminf_{t\to 0} |T_t f(x)| > \lambda\right\}\right) \\
&\le \mu\left(\left\{x \in X : \limsup_{t\to 0} |T_t (f - f_n)(x)| - \liminf_{t\to 0} |T_t (f - f_n)(x)| > \lambda\right\}\right) \\
&\le 2\mu\left(\left\{x \in X : T^*(f - f_n)(x) > \frac{\lambda}{2}\right\}\right) \\
&\le 2\left(\frac{2c}{\lambda}\|f - f_n\|_{L^p}\right)^q,
\end{aligned}$$

as T^* is weak-(p,q). As $n \to \infty$, the bottom line tends to zero, hence the top line must equal zero, and by the same argument as before we obtain the result. $\square$

We now define the *Hardy–Littlewood maximal function.* We will use this maximal function to bound the maximal operator associated with an approximation to the identity, and hence to prove almost everywhere convergence in the Cesàro, Abel–Poisson and Gauß–Weierstraß senses.

Definition 2.38. *Given $f \in L^1(\mathbb{R}^n)$, define the* Hardy–Littlewood maximal function *by*

$$\begin{aligned} Mf(x) &:= \sup_{r>0} \frac{1}{\mathrm{vol}(B_r(0))} \int_{B_r(0)} |f(x-y)| \,\mathrm{d}y \\ &= \sup_{r>0} \frac{1}{\mathrm{vol}(B_r(x))} \int_{B_r(x)} |f(y)| \,\mathrm{d}y. \end{aligned}$$

Why balls, and not cubes? We could quite happily define

$$\begin{aligned} M_{\mathrm{Q}}f(x) &:= \sup_{r>0} \frac{1}{\mathrm{vol}(Q_r(0))} \int_{Q_r(0)} |f(x-y)| \,\mathrm{d}y \\ &= \sup_{r>0} \frac{1}{\mathrm{vol}(Q_r(x))} \int_{Q_r(x)} |f(y)| \,\mathrm{d}y, \end{aligned}$$

where

$$Q_r(x) := \prod_{i=1}^{n} [x_i - r, x_i + r].$$

In fact, it is not hard to show that there are constants c, C such that

$$cM_Q f(x) \le Mf(x) \le CM_Q f(x)$$

for all f; given a square, we can inscribe and circumscribe circles, or vice versa.

In general, let R be a rectangle containing x, not necessarily centred at x, and not necessarily parallel to the axes. We may define

$$M_{\mathrm{R}}f(x) = \sup_{R} \frac{1}{\mathrm{vol}(R)} \int_R |f(y)| \,\mathrm{d}y.$$

We could now ask if

$$\lim_{\mathrm{vol}(R)\to 0} \frac{1}{\mathrm{vol}(R)} \int_R |f(y)| \,\mathrm{d}y = f(x)$$

almost everywhere: this is *false.* A counterexample is based on the Architect's Paradox. Take $Q = [0,1]\times[0,1]$ to be the unit square: then there exists a set $S \subset Q$ with $\mathrm{vol}(S) = 1$ such that for every $x \in S$ there is a line segment ℓ_x that extends from x to ∂Q such that $\ell_x \cap S = \{x\}$. (Thus you can build a hotel that fills the area of Q where you can always see the ocean.)

We now use the Hardy–Littlewood maximal function to bound the maximal operator associated to an approximation to the identity. This will lead to a proof that almost everywhere convergence in the Cesàro, Abel–Poisson and Gauß–Weierstraß senses holds. First, recall that, given $\phi \in L^1(\mathbb{R}^n)$, we set $\phi_t(x) := \frac{1}{t^n}\phi(\frac{x}{t})$; and then,

$$\phi_t * f(x) = \int_{\mathbb{R}^n} \frac{1}{t^n} \phi\left(\frac{y}{t}\right) f(x-y) \,\mathrm{d}y.$$

Theorem 2.39. *Let $\phi\colon \mathbb{R}^n \to \mathbb{R}$ be a non-negative radial function in $L^1(\mathbb{R}^n)$ which is decreasing as a function of radius, i.e. if $|x| \leq |y|$ then $\phi(x) \geq \phi(y)$. Then, for all $x \in \mathbb{R}^n$,*

$$\sup_{t>0} |\phi_t * f(x)| \leq \|\phi\|_{L^1} Mf(x).$$

Proof. First, assume that ϕ is a simple function. Then it can be written in the form

$$\phi(x) = \sum_{j=1}^{k} a_j \chi_{B_{R_j}}(x),$$

where $B_{R_j} := B_{R_j}(0)$, the coefficients $a_j > 0$, and the radii $R_1 \geq R_2 \geq \cdots \geq R_k$. Then for any $f \in L^1(\mathbb{R}^n)$,

$$\begin{aligned}
\phi * f(x) &= \int_{\mathbb{R}^n} \sum_{j=1}^{k} a_j \chi_{B_{R_j}}(y) f(x-y)\,\mathrm{d}y \\
&= \sum_{j=1}^{k} a_j \operatorname{vol}(B_{R_j}) \cdot \frac{1}{\operatorname{vol}(B_{R_j})} \int_{\mathbb{R}^n} \chi_{B_{R_j}}(y) f(x-y)\,\mathrm{d}y \\
&= \sum_{j=1}^{k} a_j \operatorname{vol}(B_{R_j}) \cdot \frac{1}{\operatorname{vol}(B_{R_j})} \int_{B_{R_j}} f(x-y)\,\mathrm{d}y \\
&\leq \sum_{j=1}^{k} a_j \operatorname{vol}(B_{R_j}) \cdot \sup_{1\leq j\leq k} \frac{1}{\operatorname{vol}(B_{R_j})} \int_{B_{R_j}} f(x-y)\,\mathrm{d}y \\
&\leq \sum_{j=1}^{k} a_j \operatorname{vol}(B_{R_j}) \cdot Mf(x) \\
&= \int_{\mathbb{R}^n} \phi(y)\,\mathrm{d}y \cdot Mf(x),
\end{aligned}$$

so $\phi * f(x) \leq \|\phi\|_{L^1} Mf(x)$. Since ϕ_t is another non-negative radially decreasing function, and

$$\int_{\mathbb{R}^n} \frac{1}{t^n} \phi\left(\frac{y}{t}\right) = \int_{\mathbb{R}^n} \phi(y)\,\mathrm{d}y,$$

we have that

$$\phi_t * f(x) \leq \|\phi_t\|_{L^1} Mf(x) = \|\phi\|_{L^1} Mf(x),$$

and hence

$$\sup_{t>0} |\phi_t * f(x)| \leq \|\phi\|_{L^1} Mf(x)$$

for all simple ϕ satisfying the assumptions.

For a general function ϕ satisfying the assumptions, let ϕ_n be an increasing sequence of simple functions which also satisfy the assumptions such that $\phi_n \to \phi$. Then, by the monotone convergence theorem, we have that

$$|\phi_t * f(x)| = \lim_{n\to\infty} |(\phi_n)_t * f(x)| \leq \lim_{n\to\infty} \|\phi_n\|_{L^1} Mf(x) = \|\phi\|_{L^1} Mf(x),$$

as required. □

Corollary 2.40. *Given $\phi \in L^1(\mathbb{R}^n)$, let $\psi \in L^1(\mathbb{R}^n)$ be a non-negative radially decreasing function such that $|\phi(x)| \le \psi(x)$ for all $x \in \mathbb{R}^n$. Then, for all $x \in \mathbb{R}^n$,*

$$\sup_{t>0} |\phi_t * f(x)| \le \|\psi\|_{L^1} Mf(x).$$

Proof. We observe that

$$\phi_t * f(x) = \int_{\mathbb{R}^n} \frac{1}{t^n}\phi\left(\frac{y}{t}\right) f(x-y)\,\mathrm{d}y \le \int_{\mathbb{R}^n} \frac{1}{t^n}\psi\left(\frac{y}{t}\right) f(x-y)\,\mathrm{d}y = \psi_t * f(x),$$

and the result follows immediately. □

We now wish to combine theorem 2.36 with corollary 2.40. We first need a covering lemma, due to Vitali:

Lemma 2.41 (Vitali covering lemma)**.** *Let $E \subset \mathbb{R}^n$ be measurable, and let E be covered by a family of balls $\mathcal{B} := \{B_j\}_{j\in\Lambda}$ of bounded diameter; i.e.,*

$$\sup_{j\in\Lambda} \operatorname{diam} B_j < \infty.$$

Then there exists an at most countable subfamily $\{\overline{B_k}\}_{k=1}^{\infty}$ of $\{B_j\}_{j\in\Lambda}$ that are pairwise disjoint (i.e. $\overline{B_i} \cap \overline{B_j} = \varnothing$ if $i \ne j$), and

$$\sum_{k=1}^{\infty} \operatorname{vol}(\overline{B_k}) \ge \frac{1}{5^n}\operatorname{vol}(E).$$

Proof. Given a collection $\mathcal{B} := \{B_j\}_{j\in\Lambda}$ with $M := \sup\{\operatorname{diam} B_j : B_j \in \mathcal{B}\} < \infty$, we choose the subfamily $\{\overline{B_k}\}_{k=1}^{\infty}$ inductively, as follows. First, pick $\overline{B_1}$ to be any ball in $\mathcal{B}$ such that $\operatorname{diam}(\overline{B_1}) > \frac{1}{2}M$. Then, assuming that $\overline{B_1}, \dots, \overline{B_k}$ have been chosen, define

$$\mathcal{B}_k := \{B_j \in \mathcal{B} : B_j \cap \overline{B_i} = \varnothing \text{ for all } 1 \le i \le k\},$$

and

$$M_k := \sup\{\operatorname{diam} B_j : B_j \in \mathcal{B}_k\}.$$

If $\mathcal{B}_k$ is nonempty, we then choose $\overline{B_{k+1}}$ to be any ball in $\mathcal{B}_k$ such that $\operatorname{diam}\overline{B_{k+1}} > \frac{1}{2}M_k$. If $\mathcal{B}_k = \varnothing$, then we stop and we have a finite subfamily; otherwise, we have a countable subfamily. For simplicity of notation, if $\mathcal{B}_k = \varnothing$ then we simply set $\overline{B_\ell} := \varnothing$ for $\ell > k$. If $\sum_{k=1}^{\infty}\operatorname{vol}(\overline{B_k}) = \infty$, then certainly $\sum_{k=1}^{\infty}\operatorname{vol}(\overline{B_k}) \ge \frac{1}{5^n}\operatorname{vol}(E)$. So suppose that $\sum_{k=1}^{\infty}\operatorname{vol}(\overline{B_k}) < \infty$.

We now define $\overline{B_k^*}$ to be the ball with the same centre as $\overline{B_k}$ and 5 times the radius. We claim that $E \subseteq \bigcup_{k=1}^{\infty}\overline{B_k^*}$. Pick $x \in E$. As $\mathcal{B}$ is a cover of E, there exists a $B_j \in \mathcal{B}$ such that $x \in B_j$. As $\sum_{k=1}^{\infty}\operatorname{vol}(\overline{B_k}) < \infty$, we have that $\operatorname{diam}\overline{B_k} \to 0$ as $k \to \infty$. So choose K to be the smallest natural number such that

$$\operatorname{diam}\overline{B_{K+1}} < \frac{1}{2}\operatorname{diam} B_j;$$

i.e. such that $\operatorname{diam} \overline{B_k} \geq \frac{1}{2} \operatorname{diam} B_j$ for $1 \leq k \leq K$. Then B_j must intersect with at least one of $\overline{B_1}, \dots, \overline{B_K}$, since if B_j were disjoint from $\overline{B_1}, \dots, \overline{B_K}$, we would have $B_j \in \mathcal{B}_K$, and then

$$\operatorname{diam} \overline{B_{K+1}} > \frac{1}{2} M_k \geq \frac{1}{2} \operatorname{diam} B_j,$$

which is a contradiction. So choose $1 \leq K_0 \leq K$ such that $B_j \cap \overline{B_{K_0}} \neq \varnothing$. Since $\operatorname{diam} B_j \leq 2 \operatorname{diam} \overline{B_{K_0}} = 4 \operatorname{radius} \overline{B_{K_0}}$, expanding $\overline{B_{K_0}}$ to five times its radius will ensure that it contains B_j; that is, $B_j \subseteq \overline{B^*_{K_0}}$.

Hence $E \subseteq \bigcup_{k=1}^\infty \overline{B^*_k}$, and so

$$\operatorname{vol}(E) \leq \operatorname{vol}\left(\bigcup_{k=1}^\infty \overline{B^*_k}\right) \leq \sum_{k=1}^\infty \operatorname{vol}(\overline{B^*_k}) = 5^n \sum_{k=1}^\infty \operatorname{vol}(\overline{B_k}),$$

so that $\sum_{k=1}^\infty \operatorname{vol}(\overline{B_k}) \geq \frac{1}{5^n} \operatorname{vol}(E)$, as required. $\square$

We can now prove that the Hardy–Littlewood maximal function M is weak-$(1,1)$ and weak-(∞,∞), which by the Marcinkiewicz interpolation theorem will prove that M is strong-(p,p) for all $1 < p < \infty$. We will do so using the Vitali covering lemma.

Theorem 2.42. *The Hardy–Littlewood maximal function M is weak-$(1,1)$ and strong-(p,p) for all $1 < p < \infty$.*

Proof. First, observe that

$$Mf(x) = \sup_{r>0} \frac{1}{\operatorname{vol}(B_r(x))} \int_{B_r(x)} |f(y)| \, \mathrm{d}y \leq \|f\|_{L^\infty},$$

so that $\|Mf\|_{L^\infty} \leq \|f\|_{L^\infty}$, and hence M is weak-(∞,∞). Thus by the Marcinkiewicz interpolation theorem, it suffices to prove that M is weak-$(1,1)$, i.e. that

$$\mu(\{x \in \mathbb{R}^n : Mf(x) > \lambda\}) \leq \frac{c\|f\|_{L^1}}{\lambda}$$

for all $f \in L^1(\mathbb{R}^n)$ and all $\lambda > 0$.

Fix $\lambda > 0$, and define $E_\lambda := \{x \in \mathbb{R}^n : Mf(x) > \lambda\}$. If $x \in E_\lambda$, then

$$Mf(x) = \sup_{r>0} \frac{1}{\operatorname{vol}(B_r(x))} \int_{B_r(x)} |f(y)| \, \mathrm{d}y > \lambda;$$

hence for each $x \in E_\lambda$ there exists $R_x > 0$ such that

$$\frac{1}{\operatorname{vol}(B_{R_x}(x))} \int_{B_{R_x}(x)} |f(y)| \, \mathrm{d}y \geq \lambda.$$

Thus E_λ is covered by balls $\{B_{R_x}(x) : x \in E_\lambda\}$, with

$$\operatorname{vol}(B_{R_x}(x)) \leq \frac{1}{\lambda} \int_{B_{R_x}(x)} |f(y)| \, \mathrm{d}y.$$

Suppose that $\{R_x : x \in E_\lambda\}$ is not bounded above: then there exists a sequence $x_j \in E_\lambda$ such that $R_{x_j} \to +\infty$; but then

$$\int_{\mathbb{R}^n} |f(y)|\,\mathrm{d}y \geq \int_{B_{R_{x_j}}} |f(y)|\,\mathrm{d}y \geq \lambda\,\mathrm{vol}(B_{R_{x_j}}) \to \infty$$

as $j \to \infty$, since λ is fixed; and then f would not be in $L^1(\mathbb{R}^n)$.

Hence the collection $\{B_{R_x}(x) : x \in E_\lambda\}$ is a covering of E_λ by balls of bounded diameter, so by the Vitali covering lemma there exists a disjoint subfamily $\{\overline{B_j}\}_{j=1}^{\infty}$ such that

$$\frac{1}{5^n}\mu(E_\lambda) \leq \sum_{j=1}^{\infty} \mathrm{vol}(\overline{B_j}).$$

Hence

$$\mu(E_\lambda) \leq 5^n \sum_{j=1}^{\infty} \mathrm{vol}(\overline{B_j}) \leq \frac{5^n}{\lambda} \sum_{j=1}^{\infty} \int_{\overline{B_j}} |f(y)|\,\mathrm{d}y = \frac{5^n}{\lambda} \int_{\bigcup_{j=1}^{\infty} \overline{B_j}} |f(y)|\,\mathrm{d}y \leq \frac{5^n \|f\|_{L^1}}{\lambda},$$

so that M is weak-$(1,1)$. This completes the proof. □

We can now use this to show almost everywhere convergence of certain approximations to the identity:

Corollary 2.43. *Let $\phi \in L^1(\mathbb{R}^n)$ such that there exists $\psi \in L^1(\mathbb{R}^n)$ which is non-negative, radially decreasing, and $|\phi(x)| \leq \psi(x)$ for all $x \in \mathbb{R}^n$. Fix $1 \leq p < \infty$, and let $f \in L^p(\mathbb{R}^n)$. Then*

$$\lim_{t \to 0} \phi_t * f(x) = \left(\int_{\mathbb{R}^n} \phi(y)\,\mathrm{d}y \right) f(x)$$

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

Proof. Suppose that $\int_{\mathbb{R}^n} \phi(y)\,\mathrm{d}y = 1$ (if not, replace ϕ by $\frac{\phi}{\left(\int_{\mathbb{R}^n} \phi(y)\,\mathrm{d}y\right)}$). Define $T_t f(x) := \phi_t * f(x)$, and denote by

$$T^* f(x) := \sup_{t>0} |T_t f(x)| = \sup_{t>0} |\phi_t * f(x)|$$

the maximal operator associated to the family $\{T_t\}_{t>0}$. We know from corollary 2.40 that

$$T^* f(x) \leq \|\psi\|_{L^1} M f(x).$$

By theorem 2.42, M is weak-(p,p) for $1 \leq p < \infty$, so

$$\mu(\{x \in X : |T^* f(x)| > \lambda\}) \leq \mu\left(\left\{ x \in X : |Mf(x)| > \frac{\lambda}{\|\psi\|_{L^1}} \right\} \right) \leq \left(\frac{c\|\psi\|_{L^1} \|f\|_{L^p}}{\lambda} \right)^p,$$

and hence T^* is weak-(p,p) for $1 \leq p < \infty$. Hence by theorem 2.36, the set

$$E := \{f \in L^p(\mathbb{R}^n) : \lim_{t \to 0} T_t f(x) = f(x) \text{ for a.e. } x \in X\}$$

is closed in $L^p(\mathbb{R}^n)$. Now, by theorem 2.25, if $f \in \mathcal{S}(\mathbb{R}^n)$ then $T_t f(x) \to f(x)$ uniformly over all $x \in \mathbb{R}^n$, and so in particular it converges almost everywhere; hence $\mathcal{S}(\mathbb{R}^n) \subset E$. As $\mathcal{S}(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$, $E = L^p(\mathbb{R}^n)$; i.e., for every $f \in L^p(\mathbb{R}^n)$, $T_t f(x) \to f(x)$ for almost every $x \in \mathbb{R}^n$. □

Finally, we now apply this corollary to obtain convergence of the Fourier transform in the Cesàro, Abel–Poisson and Gauß–Weierstraß senses.

Cesàro convergence: We have $\sigma_R f = F_R * f$, where $F_R(x) = \frac{\sin^2(\pi R x)}{R(\pi x)^2}$ in dimension 1. We now prove that $F_R * f$ converges to f almost everywhere:

Proposition 2.44. *Let $1 \le p < \infty$. For any $f \in L^p(\mathbb{R})$, we have that $F_R * f(x) \to f(x)$ for almost every $x \in \mathbb{R}$ as $R \to +\infty$.*

Proof. As in proposition 2.27, F_R is an approximation to the identity, with

$$\phi(x) := F_1(x) = \frac{\sin^2 \pi x}{(\pi x)^2}.$$

Setting $\psi(x) = \min\{1, \frac{1}{(\pi x)^2}\}$, we see that ψ is non-negative and decreasing, and that $|\phi(x)| \le \psi(x)$ for every x. Hence by corollary 2.43, $F_{1/t} * f(x) = \phi_t * f \to f$ almost everywhere for any $f \in L^p(\mathbb{R})$. $\square$

Abel–Poisson convergence: We have

$$u(x,t) = \int_{\mathbb{R}^n} e^{-2\pi t|\xi|} \hat{f}(\xi) e^{2\pi i x\cdot\xi} \, \mathrm{d}\xi = P_t * f(x),$$

where

$$P_t(x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{t}{(t^2 + |x|^2)^{(n+1)/2}}.$$

We now prove that $P_t * f$ converges to f almost everywhere:

Proposition 2.45. *Let $1 \le p < \infty$. For any $f \in L^p(\mathbb{R}^n)$, we have that $P_t * f(x) \to f(x)$ for almost every $x \in \mathbb{R}^n$ as $t \to 0^+$.*

Proof. As in proposition 2.28, P_t is an approximation to the identity with

$$\phi(x) := P_1(x) = \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \frac{1}{(1 + |x|^2)^{(n+1)/2}}.$$

As ϕ is non-negative, radial and decreasing, by corollary 2.43 $P_t * f \to f$ almost everywhere for any $f \in L^p(\mathbb{R}^n)$. $\square$

Gauß–Weierstraß convergence: We have

$$\omega(x,t) = \int_{\mathbb{R}^n} e^{-4\pi^2 t|\xi|^2} \hat{f}(\xi) e^{2\pi i x\cdot\xi} \, \mathrm{d}\xi = W_t * f(x),$$

where

$$W_t(x) = \frac{1}{(4\pi t)^{n/2}} e^{-|x|^2/4t}.$$

We now prove that $W_t * f$ converges to f almost everywhere:

Proposition 2.46. *Let $1 \le p < \infty$. For any $f \in L^p(\mathbb{R}^n)$, we have that $W_t * f(x) \to f(x)$ for almost every $x \in \mathbb{R}^n$ as $t \to 0^+$.*

Proof. As in proposition 2.29, W_{t^2} is an approximation to the identity with

$$\phi(x) := W_1(x) = \frac{1}{(4\pi)^{n/2}} e^{-|x|^2/4}.$$

As ϕ is non-negative, radial and decreasing, by corollary 2.43 $W_{t^2} * f \to f$ almost everywhere for any $f \in L^p(\mathbb{R}^n)$, and hence $W_t * f \to f$ almost everywhere. $\square$

2.8 The Lebesgue Differentiation Theorem*

The question which motivated the Hardy–Littlewood maximal function was that eventually proved in Lebesgue's differentiation theorem: does

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

for almost every $x \in \mathbb{R}^n$? If f is continuous this is obvious, but if f is only in L^1 then it is nontrivial.

Theorem 2.47 (Lebesgue's differentiation theorem)**.** *If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then*

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

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

Proof. Define the family of operators $\{M_r\}_{r>0}$ by

$$M_r f(x) := \frac{1}{\operatorname{vol}(B_r(x))}\int_{B_r(x)}|f(y)|\,\mathrm{d}y.$$

The maximal operator associated to the family $\{M_r\}_{r>0}$ is precisely the Hardy–Littlewood maximal function M, which by theorem 2.42 is weak-$(1,1)$. Hence by theorem 2.36, the set

$$E := \{f \in L^1(\mathbb{R}^n) : \lim_{r\to 0} M_r f(x) = f(x) \text{ for a.e. } x \in X\}$$

is closed in $L^1(\mathbb{R}^n)$.

Now, if $f \in C^0(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, then given $\varepsilon > 0$ there exists $\delta > 0$ such that if $|y - x| < \delta$ then $|f(y) - f(x)| < \varepsilon$, and hence

$$\frac{1}{\operatorname{vol}(B_r(x))}\int_{B_r(x)}|f(y) - f(x)|\,\mathrm{d}y < \varepsilon$$

whenever $|y - x| < \delta$. Hence $C^0(\mathbb{R}^n) \cap L^1(\mathbb{R}^n) \subset E$, and as $C^0(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$, $E = L^1(\mathbb{R}^n)$.

If $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$ then given a compact subset $K \subset \mathbb{R}^n$ we have that $f\chi_K \in L^1(\mathbb{R}^n)$, and hence that the result holds for almost every $x \in K$. Hence, by taking a covering of $\mathbb{R}^n$ by compact sets — e.g., consider the cover $\{K_m\}_{m\in\mathbb{Z}^n}$ given by

$$K_m := [m_1, m_1 + 1] \times \cdots \times [m_n, m_n + 1]$$

— we see that the result holds for almost every $x \in \mathbb{R}^n$. □

In fact, the same proof applies to the stronger result that, if $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n)$, then

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

for almost every $x \in \mathbb{R}^n$, since the theorem implies that $|f(x)| \leq Mf(x)$ for almost every $x \in \mathbb{R}^n$, and hence that

$$\sup_{r>0} \frac{1}{\operatorname{vol}(B_r(x))} \int_{B_r(x)} |f(y) - f(x)| \, \mathrm{d}y \leq Mf(x) + |f(x)| \leq 2Mf(x),$$

so as M is weak-$(1,1)$, so is the left-hand side, and hence the result follows.

It may be remarked that, while M is weak-$(1,1)$ and strong-(p,p), it is *not* strong-$(1,1)$. Indeed, the strong-$(1,1)$ inequality never holds, as shown by the following result:

Proposition 2.48. *If $f \in L^1(\mathbb{R}^n)$ and f is not identically zero, then $Mf \notin L^1(\mathbb{R}^n)$.*

Proof. Let $f \in L^1(\mathbb{R}^n)$ be not identically zero; then there must exist $R > 0$ such that

$$\int_{B_R(0)} |f| \geq \varepsilon > 0.$$

Then if $|x| > R$, then $B_R(0) \subset B_{2|x|}(x)$, and hence

$$\begin{aligned} Mf(x) &= \sup_{r>0} \frac{1}{\operatorname{vol}(B_r(x))} \int_{B_r(x)} |f(y)| \, \mathrm{d}y \\ &\geq \frac{1}{\operatorname{vol}(B_{2|x|}(x))} \int_{B_{2|x|}(x)} |f(y)| \, \mathrm{d}y \\ &\geq \frac{1}{\operatorname{vol}(B_{2|x|}(x))} \int_{B_R(0)} |f(y)| \, \mathrm{d}y \\ &> \frac{\Gamma(\frac{n}{2}+1)}{\pi^{n/2}} \cdot \frac{\varepsilon}{2^n |x|^n}, \end{aligned}$$

and $x \mapsto \frac{1}{|x|^n}$ is not an $L^1(\mathbb{R}^n)$ function. $\square$

3 Distribution Theory

In order to utilise the Fourier transform in its greatest possible generality, we define "generalised functions" — although that is something of a misnomer. The correct "generalisation" of the notion of function is called a *distribution*, which is (in essence) a linear functional on a suitable space of test functions. In order to motivate the generalisation, we first consider *weak derivatives*. We remind the reader of the notation introduced in section 2.2, which will be used heavily throughout.

3.1 Weak Derivatives

Consider an open subset $\Omega \subseteq \mathbb{R}^n$ such that the boundary $\partial\Omega$ is C^1. If $\phi \in C_c^\infty(\Omega)$ and $u \in C^k(\Omega)$, then we may integrate by parts: for any multi-index α with $|\alpha| \leq k$, we have

$$\int_\Omega \partial^\alpha u(x)\phi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_\Omega u(x)\partial^\alpha\phi(x)\,\mathrm{d}x,$$

where the boundary terms are zero because ϕ has compact support in Ω. This is true for every $\phi \in C_c^\infty(\Omega)$.

Definition 3.1 (Weak derivative)**.** *Consider an open subset $\Omega \subseteq \mathbb{R}^n$ such that the boundary $\partial\Omega$ is C^1. Let $u, v \in L^1_{\mathrm{loc}}(\Omega)$, and let $\alpha \in \mathbb{N}_0^n$ be a multi-index. We say that v is the* α^{th} weak derivative of *u if, for every $\phi \in C_c^\infty(\Omega)$, we have*

$$\int_\Omega v(x)\phi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_\Omega u(x)\partial^\alpha\phi(x)\,\mathrm{d}x.$$

We write $\partial^\alpha u := v$.

We shall use the following lemma without proof:

Lemma 3.2. *Let $\Omega \subseteq \mathbb{R}^n$ is open, and let $f \in L^1_{\mathrm{loc}}(\Omega)$. If $\int_\Omega f(x)\phi(x)\,\mathrm{d}x = 0$ for all $\phi \in C_c^\infty(\Omega)$, then $f(x) = 0$ for almost every $x \in \Omega$.*

Lemma 3.3. *Weak derivatives, when they exist, are unique almost everywhere.*

Proof. Let v_1 and v_2 be weak derivatives of u, i.e.

$$\int_\Omega u(x)\partial^\alpha\phi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_\Omega v_1(x)\phi(x)\,\mathrm{d}x = (-1)^{|\alpha|}\int_\Omega v_2(x)\phi(x)\,\mathrm{d}x$$

for all $\phi \in C_c^\infty(\Omega)$. Then $\int_\Omega v_1(x)\phi(x)\,\mathrm{d}x = \int_\Omega v_2(x)\phi(x)\,\mathrm{d}x$ for all $\phi \in C_c^\infty(\Omega)$, so that $\int_\Omega(v_1(x) - v_2(x))\phi(x)\,\mathrm{d}x = 0$ for all $\phi \in C_c^\infty(\Omega)$. Hence, by lemma 3.2, $v_1 - v_2 = 0$ almost everywhere, i.e. $v_1 = v_2$ almost everywhere. $\square$

Of course, if $u \in C^k$, and $|\alpha| \leq k$, then the α^{th} weak derivative of u is just its classical derivative. But there are functions u which are *not* classically differentiable which have weak derivatives. For example, let us consider $u\colon \mathbb{R} \to \mathbb{R}$ given by $u(x) = |x|$. We claim that $v\colon \mathbb{R} \to \mathbb{R}$ given by

$$v(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

is the weak derivative of u. To see this, we split the integral and then integrate by parts:

$$\begin{aligned}\int_{-\infty}^{\infty} u(x)\phi'(x)\,\mathrm{d}x &= \int_{-\infty}^{0} -x\phi'(x)\,\mathrm{d}x + \int_{0}^{\infty} x\phi'(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{0} \phi(x)\,\mathrm{d}x - x\phi(x)|_{-\infty}^{0} - \int_{0}^{\infty} \phi(x)\,\mathrm{d}x + x\phi(x)|_{0}^{\infty} \\ &= -\int_{-\infty}^{\infty} v(x)\phi(x)\,\mathrm{d}x.\end{aligned}$$

Since this holds for any $\phi \in C_c^\infty(\mathbb{R})$, we have that $u' = v$ weakly.

However, not all functions have weak derivatives. We show now that the above v has no weak derivative, by contradiction. Suppose that w is the weak derivative of v; then it must satisfy

$$\int_{-\infty}^{\infty} w(x)\phi(x)\,\mathrm{d}x = -\int_{-\infty}^{\infty} v(x)\phi'(x)\,\mathrm{d}x = \int_{-\infty}^{0} \phi'(x)\,\mathrm{d}x - \int_{0}^{\infty} \phi'(x)\,\mathrm{d}x = 2\phi(0).$$

for every $\phi \in C_c^\infty(\mathbb{R})$. However, given $\varepsilon > 0$, if we pick $\phi \in C_c^\infty(\mathbb{R}\setminus[-\varepsilon,\varepsilon])$ then $\phi(0) = 0$, so

$$\int_{-\infty}^{\infty} w(x)\phi(x)\,\mathrm{d}x = 0$$

for all $\phi \in C_c^\infty(\mathbb{R}\setminus[-\varepsilon,\varepsilon])$. But then by lemma 3.2 $w(x) = 0$ for almost every $x \in \mathbb{R}\setminus[-\varepsilon,\varepsilon]$, and since $\varepsilon > 0$ was arbitrary, we see that $w(x) = 0$ for almost every $x \in \mathbb{R}$. Thus if $\phi \in C_c^\infty(\mathbb{R})$ has $\phi(0) \neq 0$, then

$$\int_{-\infty}^{\infty} w(x)\phi(x)\,\mathrm{d}x = 0 \neq 2\phi(0).$$

Hence v cannot have a weak derivative!

One of the primary motivations behind studying weak derivatives is solutions to PDEs. Consider, as a simple example, the transport equation

$$\begin{cases} \partial_t u(x,t) + \partial_x u(x,t) = 0 \\ \qquad\qquad\quad u(x,0) = f(x) \end{cases} \tag{8}$$

for $x, t \in \mathbb{R}$. From the equation, we see that $u(x,t) = g(x-t)$ for some function g of one variable; by the initial condition, we see that $u(x,t) = f(x-t)$. However, what if $f \notin C^1(\mathbb{R})$? Then $u(x,t) = f(x-t)$ makes sense, but it may not "solve" (8). However, formally, given any $\phi \in C_c^\infty(\mathbb{R}^2)$, we have that

$$0 = \iint_{\mathbb{R}^2} (\partial_t u + \partial_x u)\phi\,\mathrm{d}x\,\mathrm{d}t = -\iint_{\mathbb{R}^2} u(\partial_t\phi + \partial_x\phi)\,\mathrm{d}x\,\mathrm{d}t$$

So, we say that u solves (8) *weakly* if $u(x,0) = f(x)$ for all $x \in \mathbb{R}$ and

$$\iint_{\mathbb{R}^2} u(\partial_t\phi + \partial_x\phi)\,\mathrm{d}x\,\mathrm{d}t = 0$$

for all $\phi \in C_c^\infty(\mathbb{R}^2)$. So, if f has a weak derivative, then $u(x,t) = f(x-t)$ is a *weak* solution of (8).

3.2 Distributions: Basic Definitions

ADD STUFF ON FUNDAMENTAL SOLUTION OF LAPLACIAN?

If X is a Banach space, and $u\colon X \to \mathbb{R}$ is a linear functional on X, then we write $\langle u, \phi\rangle := u(\phi)$, to represent the pairing between X and its dual space X^*.

For $1 < p < \infty$, the dual of $L^p(X)$ is $L^q(X)$, with $1/p + 1/q = 1$. That is, to each $f \in L^q(X)$ there corresponds a unique bounded linear functional $\phi\colon L^p(X) \to \mathbb{C}$ such that $\langle \phi, g\rangle = \int_X fg \,\mathrm{d}\mu$ for all $g \in L^p(X)$.

It turns out that the "correct" way to generalise the notion of function is by considering the dual space of $C_c^\infty(X)$. However, $C_c^\infty(X)$ is not a Banach space, so we must come up with an alternative definition of the dual space.

Definition 3.4 (Distribution)**.** *Let $X \subset \mathbb{R}^n$ be open, and let $u\colon C_c^\infty(X) \to \mathbb{C}$ be a linear functional. u is called a* distribution *on X if, for every compact subset $K \subset X$, there exists a constant $C > 0$ and a natural number N such that*

$$|\langle u, \phi\rangle| \leq C \sum_{|\alpha|\leq N} \sup_{x\in K} |\partial^\alpha \phi| \tag{9}$$

for all $\phi \in C_c^\infty(K)$. The set of all distributions on X is denoted by $\mathcal{D}'(X)$.

Note carefully that C and N may depend on K. Since $C_c^\infty(X)$ is not a Banach space, we need (9) to encapsulate the boundedness of the linear functional X. It would be nice if our notion of "generalised function" really were a generalisation of the notion of a function, so we show now that each function $f \in L^p_{\mathrm{loc}}(X)$ defines a distribution in $\mathcal{D}'(X)$.

Proposition 3.5. *Let $X \subset \mathbb{R}^n$ be open, let $1 \leq p \leq \infty$, and let $f \in L^p_{\mathrm{loc}}(X)$. Then T_f, defined by $\langle T_f, \phi\rangle := \int_X f(x)\phi(x)\,\mathrm{d}x$ is a distribution in $\mathcal{D}'(X)$.*

Proof. It is clear that T_f is a linear functional on $C_c^\infty(X)$; we must prove that (9) holds. Pick a compact subset $K \subset X$, and let $\phi \in C_c^\infty(K)$. Then, by Hölder's inequality, we have

$$|\langle T_f, \phi\rangle| = \left|\int_K f(x)\phi(x)\,\mathrm{d}x\right| \leq \|f\|_{L^p(K)}\|\phi\|_{L^q(K)} \leq \|f\|_{L^p(K)}(\mu(K))^{1/q} \sup_{x\in K}|\phi(x)|.$$

(Here we use the convention that $1/q = 0$ when $q = \infty$.) As $f \in L^p_{\mathrm{loc}}(X)$, we have that $C := \|f\|_{L^p(K)}(\mu(K))^{1/q} < \infty$. So taking $N = 0$, we see that T_f is a distribution. $\square$

In particular, this works for every continuous function $f \in C^0(X)$. We will often — though not always — abuse notation and use f to mean the distribution T_f.

A very important example of a distribution which does *not* come from a function is the *Dirac delta function*, which is not a function at all but rather a distribution. We define $\delta\colon C_c^\infty(\mathbb{R}^n) \to \mathbb{C}$ by

$$\langle \delta, \phi\rangle := \phi(0).$$

δ is clearly a linear functional; to see that it is a distribution, observe that for any $\phi \in C_c^\infty(K)$, we have

$$|\langle \delta, \phi\rangle| = |\phi(0)| \leq \sup_{x\in K}|\phi(x)|,$$

so we may take $C = 1$ and $N = 0$ in (9). Similarly, we may define $\delta_y \colon C_c^\infty(\mathbb{R}^n) \to \mathbb{C}$ by

$$\langle \delta_y, \phi \rangle := \phi(y).$$

Again, this is a distribution.

Since $C_c^\infty(X)$ is not a Banach space, we do not, as yet, have a notion of convergence in it (and hence no topology). We now define a notion of convergence in $C_c^\infty(X)$ such that, with respect to the topology so generated, $\mathcal{D}'(X)$ is indeed the topological dual of $C_c^\infty(X)$.

Definition 3.6 (Convergence in C_c^∞). *Let $X \subset \mathbb{R}^n$ be open. We say that a sequence $(\phi_j)_{j=1}^\infty$ in $C_c^\infty(X)$ converges to $\phi \in C_c^\infty(X)$ if there exists a compact set $K \subset X$ such that $\operatorname{spt}(\phi_j) \subset K$ for all j, $\operatorname{spt}(\phi) \subset K$, and for every multi-index α we have that*

$$\sup_{x \in K} |\partial^\alpha \phi_j(x) - \partial^\alpha \phi(x)| \to 0$$

as $j \to \infty$ (that is, $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly on K).

With respect to this notion of convergence, $\mathcal{D}'(X)$ is indeed the dual of $C_c^\infty(X)$:

Theorem 3.7. *A linear functional $u \colon C_c^\infty(X) \to \mathbb{C}$ is a distribution if, and only if, for every sequence $(\phi_j)_{j=1}^\infty$ in $C_c^\infty(X)$ such that $\phi_j \to \phi$ with $\phi \in C_c^\infty(X)$, we have that*

$$\lim_{j \to \infty} \langle u, \phi_j \rangle = \langle u, \phi \rangle.$$

Before we proceed to the proof, note that $\phi_j \to \phi$ in $C_c^\infty(X)$ if and only if $\phi_j - \phi \to 0$, so without loss of generality we may suppose that $\phi = 0$.

Proof. Suppose that u is a distribution. If $\phi_j \to 0$ in $C_c^\infty(X)$, then there is a compact set K such that $\operatorname{spt}(\phi_j) \subset K$ for every $j \in \mathbb{N}$, and $\partial^\alpha \phi_j \to 0$ uniformly in K. For this K, there exist C and N such that

$$|\langle u, \phi_j \rangle| \leq C \sum_{|\alpha| \leq N} \sup_{x \in K} |\partial^\alpha \phi_j|$$

for every $j \in \mathbb{N}$; hence $\langle u, \phi_j \rangle \to 0$ as $j \to \infty$, as required.

For the converse, suppose that $\langle u, \phi_j \rangle \to 0$ as $j \to \infty$ whenever $\phi_j \to 0$ in $C_c^\infty(X)$. Suppose, for a contradiction, that there is a compact subset $K \subset X$ such that

$$\frac{|\langle u, \phi \rangle|}{\sum_{|\alpha| \leq N} |\partial^\alpha \phi|}$$

is unbounded over $\phi \in C_c^\infty(K)$ for every N. Thus, for each $N \in \mathbb{N}$, there exists a function $\phi_N \in C_c^\infty(K)$ such that

$$\frac{|\langle u, \phi_N \rangle|}{\sum_{|\alpha| \leq N} |\partial^\alpha \phi_N|} > N.$$

for all N. Define

$$\psi_N := \frac{\phi_N}{N \sum_{|\alpha| \leq N} |\partial^\alpha \phi_N|}.$$

Since $\operatorname{spt}(\phi_N) \subset K$, we have that $\operatorname{spt}(\psi_N) \subset K$. Furthermore,

$$\sup_{x \in K} |\partial^\alpha \psi_N| = \frac{1}{N} \frac{|\partial^\alpha \phi_N|}{\sum_{|\alpha| \le N} |\partial^\alpha \phi_N|} \le \frac{1}{N} \to 0$$

as $N \to \infty$. So $\psi_N \to 0$ in $C_c^\infty(X)$, but

$$|\langle u, \psi_N \rangle| = \frac{|\langle u, \phi_N \rangle|}{N \sum_{|\alpha| \le N} |\partial^\alpha \phi_N|} > 1$$

for all n, and hence $|\langle u, \psi_N \rangle| \not\to 0$ as $N \to \infty$. This contradiction completes the proof. $\square$

Definition 3.8 (Order of a distribution)**.** *Let $X \subset \mathbb{R}^n$ be open. A distribution $u \in \mathcal{D}'(X)$ has* finite order *if there exist constants $C > 0$ and $N \in \mathbb{N}$ such that, for all compact subsets $K \subset X$, we have*

$$|\langle u, \phi \rangle| \le C \sum_{|\alpha| \le N} \sup_{x \in K} |\partial^\alpha \phi|;$$

that is, if C and N in equation (9) may be chosen independent of the compact set K. If a distribution u has finite order, we say the order *of u is the least N such that there exists $C > 0$ such that for all compact subsets $K \subset X$ we have*

$$|\langle u, \phi \rangle| \le C \sum_{|\alpha| \le N} \sup_{x \in K} |\partial^\alpha \phi|.$$

We write $\mathcal{D}'^N(X)$ to mean the space of all distributions of whose order is less than or equal to N.

For example, the delta distribution given by $\langle \delta, \phi \rangle = \phi(0)$ is of order zero. Similarly, the distribution $u \in \mathcal{D}'(\mathbb{R})$ given by $\langle u, \phi \rangle = \phi'(0)$ has

$$|\langle u, \phi \rangle| \le \sup_{x \in K} |\phi'(x)|,$$

so we may take $C = 1$ and $N = 1$ independently of K, and hence u has order 1.

Definition 3.9 (Convergence in C_c^N)**.** *Let $X \subset \mathbb{R}^n$ be open. We say that a sequence $(\phi_j)_{j=1}^\infty$ in $C_c^N(X)$ converges to $\phi \in C_c^N(X)$ if there exists a compact set $K \subset X$ such that $\operatorname{spt}(\phi_j) \subset K$ for all j, $\operatorname{spt}(\phi) \subset K$, and for every multi-index α such that $|\alpha| \le N$ we have that*

$$\sup_{x \in K} |\partial^\alpha \phi_j(x) - \partial^\alpha \phi(x)| \to 0$$

as $j \to \infty$ (that is, $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly on K).

Since $C_c^\infty(X) \subset C_c^N(X)$, we see that every sequentially continuous linear functional on $C_c^N(X)$ defines a distribution. In fact, if a distribution has finite order, the converse is true:

Proposition 3.10. *Let $X \subset \mathbb{R}^n$ be open and let $u \in \mathcal{D}'(X)$ be of finite order N. Then u has a unique extension to $C_c^N(X)$.*

This implies that the dual of $C_c^N(X)$ is precisely $\mathcal{D}'^N(X)$.

Proof. Let $u \in \mathcal{D}'(X)$ be of finite order N: then there exist $C > 0$ and $N \in \mathbb{N}$ such that for all compact subsets $K \subset X$ we have

$$|\langle u, \phi \rangle| \le C \sum_{|\alpha| \le N} \sup_{x \in K} |\partial^\alpha \phi|$$

for every $\phi \in C_c^\infty(K)$. As $C_c^\infty(K)$ is dense in $C_c^N(K)$, given $\phi \in C_c^N(K)$ we may take $\phi_j \in C_c^\infty(K)$ such that $\partial^\alpha \phi_j \to \partial^\alpha \phi$ uniformly over K for every multi-index α with $|\alpha| \le N$. Then, by the above inequality, we may define $\langle u, \phi \rangle := \lim_{j \to \infty} \langle u, \phi_j \rangle$. Thus u is uniquely defined on $C_c^N(X)$. $\square$

Definition 3.11 (Support)**.** *Let $X \subset \mathbb{R}^n$ be open and let $u \in \mathcal{D}'(X)$. The* support *of u is defined as*

$$\operatorname{spt} u := X \setminus \{x \in X : u = 0 \textit{ in a neighbourhood of } x\}.$$

Here $u = 0$ in a neighbourhood of x if there exists a compact set K such that $x \in K^o$ (that is, x lies in the interior of K) and $\langle u, \phi \rangle = 0$ for all $\phi \in C_c^\infty(K)$.

Note that as $\{x \in X : u = 0 \text{ in a neighbourhood of } x\}$ is open, $\operatorname{spt} u$ is closed. For example, the support of the delta distribution on $\mathbb{R}^n$ is $\operatorname{spt} \delta = \{0\}$, since if $x \neq 0$ we can choose a compact set K which contains x but not 0, and then for all $\phi \in C_c^\infty(K)$ we have that $\langle \delta, \phi \rangle = \phi(0) = 0$.

Consider the distribution T_f on X given by $\langle T_f, \phi \rangle := \int_X f(y)\phi(y)\,\mathrm{d}y$, where $f \in C^0(X)$. We claim that $\operatorname{spt} T_f = \operatorname{spt} f$. First, suppose that $x \in X \setminus \operatorname{spt} f$: then there exists a compact set K such that $x \in K^o$, and $f(y) = 0$ for all $y \in K$. Then $\langle T_f, \phi \rangle = \int_K f(y)\phi(y)\,\mathrm{d}y = 0$ for all $\phi \in C_c^\infty(K)$, so $x \in X \setminus \operatorname{spt} T_f$. Conversely, suppose $x \in X \setminus \operatorname{spt} T_f$: then there exists a compact set K such that $x \in K^o$, and $\langle T_f, \phi \rangle = 0$ for all $\phi \in C_c^\infty(K)$. Hence by lemma 3.2, $f(x) = 0$ for almost every $x \in K$, and as f is continuous, $f(x) = 0$ for every $x \in K$. Hence $x \in X \setminus \operatorname{spt} f$.

Definition 3.12 (Convergence of distributions)**.** *Let $X \subset \mathbb{R}^n$ be open, let $(u_j)_{j=1}^\infty$ be a sequence in $\mathcal{D}'(X)$, and let $u \in \mathcal{D}'(X)$. We say that $u_j \to u$ in $\mathcal{D}'(X)$ if*

$$\langle u_j, \phi \rangle \to \langle u, \phi \rangle$$

for every $\phi \in C_c^\infty(X)$.

For example, consider $X = (0, 2\pi)$. Then the sequence of functions $u_k(x) := e^{ikx}$ converges to 0 as $k \to \infty$ in the sense of distributions (that is, $T_{u_k} \to 0$ in $\mathcal{D}'(X)$): for all $\phi \in C_c^\infty(X)$, we have

$$\int_X e^{ikx}\phi(x)\,\mathrm{d}x \to 0$$

by the Riemann–Lebesgue lemma.

Consider a non-negative function $\rho \in C_c^\infty(\mathbb{R}^n)$ which is radial and has $\int_{\mathbb{R}^n} \rho\,\mathrm{d}x = 1$. Define $\rho_\varepsilon(x) := \frac{1}{\varepsilon^n}\rho\left(\frac{x}{\varepsilon}\right)$. Recall from lemma 2.24 that if $\phi \in C_c^\infty(\mathbb{R}^n)$ then $\rho_\varepsilon * \phi \to \phi$ pointwise as $\varepsilon \to 0$. In particular, for $x = 0$, we have

$$\int_{\mathbb{R}^n} \frac{1}{\varepsilon^n}\rho\left(\tfrac{y}{\varepsilon}\right)\phi(y)\,\mathrm{d}y \to f(0)$$

as $\varepsilon \to 0$. This says precisely that $\langle \rho_\varepsilon, \phi \rangle \to \langle \delta, \phi \rangle$ as $\varepsilon \to 0$; i.e. $\rho_\varepsilon \to \delta$ in the sense of distributions.

3.3 Distributional Derivatives and Products

Definition 3.13 (Distributional derivative)**.** *Let $X \subset \mathbb{R}^n$ be open, let $u \in \mathcal{D}'(X)$ be a distribution, and let α be a multi-index. We define the α^{th}* distributional derivative of u *to be the distribution $\partial^\alpha u$ given by*

$$\langle \partial^\alpha u, \phi \rangle = (-1)^{|\alpha|} \langle u, \partial^\alpha \phi \rangle.$$

It is clear that $\partial^\alpha u$ so defined is still a linear functional on $C_c^\infty(X)$. Since u is a distribution, for each compact subset $K \subset X$ there exists constants $C > 0$ and $N \in \mathbb{N}$ such that

$$|\langle u, \phi \rangle| \le C \sum_{|\beta| \le N} \sup_{x \in K} |\partial^\beta \phi|$$

for every $\phi \in C_c^\infty(K)$. Now,

$$|\langle \partial^\alpha u, \phi \rangle| = |\langle u, \partial^\alpha \phi \rangle| \le C \sum_{|\beta| \le N} \sup_{x \in K} |\partial^{\alpha+\beta} \phi| \le C \sum_{|\beta| \le N + |\alpha|} \sup_{x \in K} |\partial^\beta \phi|,$$

so $\partial^\alpha u$ is indeed a distribution. Thus, *every* distribution is *infinitely* differentiable!

Lemma 3.14. *Let $X \subset \mathbb{R}^n$ be open, let $(u_j)_{j=1}^\infty$ be a sequence in $\mathcal{D}'(X)$, and let $u \in \mathcal{D}'(X)$. If $u_j \to u$ in $\mathcal{D}'(X)$, then $\partial^\alpha u_j \to \partial^\alpha u$ in $\mathcal{D}'(X)$ for every multi-index α.*

Proof. We simply observe that, since $\partial^\alpha \phi$ is another test function,

$$\langle \partial^\alpha u_j, \phi \rangle = (-1)^{|\alpha|} \langle u_j, \partial^\alpha \phi \rangle \to (-1)^{|\alpha|} \langle u, \partial^\alpha \phi \rangle = \langle \partial^\alpha u, \phi \rangle. \qquad \square$$

Proposition 3.15. *Let $X \subset \mathbb{R}^n$ be open, and let $f, g \in L^1_{\text{loc}}(X)$. Then the α^{th} weak derivative $\partial^\alpha f$ exists and equals g almost everywhere if, and only if, the α^{th} distributional derivative $\partial^\alpha T_f$ equals T_g.*

Proof. Suppose the α^{th} weak derivative $\partial^\alpha f$ exists and equals g almost everywhere. Then

$$\langle \partial^\alpha T_f, \phi \rangle = (-1)^{|\alpha|} \langle T_f, \partial^\alpha \phi \rangle = (-1)^{|\alpha|} \int_X f \cdot \partial^\alpha \phi \, \mathrm{d}x = \int_X g \cdot \phi \, \mathrm{d}x = \langle T_g, \phi \rangle$$

for every $\phi \in C_c^\infty(X)$, so $\partial^\alpha T_f = T_g$. Conversely, if $\partial^\alpha T_f = T_g$ then

$$\int_X g \cdot \phi \, \mathrm{d}x = \langle T_g, \phi \rangle = \langle \partial^\alpha T_f, \phi \rangle = (-1)^{|\alpha|} \langle f, \partial^\alpha \phi \rangle = (-1)^{|\alpha|} \int_X f \cdot \partial^\alpha \phi \, \mathrm{d}x$$

for every $\phi \in C_c^\infty(X)$, so by uniqueness of weak derivatives (lemma 3.3) $\partial^\alpha f = g$ almost everywhere. $\square$

Since not every function is weakly differentiable, this tells us that if we form the distribution T_f from a function f, even if f is not infinitely differentiable, T_f will be; the price, however, is that the derivatives of T_f are represented by a function only when the weak derivatives of f exist.

Let us consider a nontrivial example of a distributional derivative. We saw earlier that the Heaviside step function $H\colon \mathbb{R} \to \mathbb{R}$ given by

$$H(x) := \begin{cases} 1 & \text{if } x > 0 \\ 1/2 & \text{if } x = 0 \\ 0 & \text{if } x < 0 \end{cases}$$

has no weak derivative, but

$$\begin{aligned} \langle \partial H, \phi \rangle &= -\langle H, \partial\phi \rangle \\ &= -\int_{\mathbb{R}} H(x)\partial\phi(x)\,\mathrm{d}x \\ &= -\int_0^\infty \partial\phi(x) \\ &= -\left.\phi(x)\right|_0^\infty \\ &= \phi(0) = \langle \delta, \phi \rangle, \end{aligned}$$

so the distributional derivative is $\partial H = \delta$. It is also easy to see that $\langle \partial^2 H, \phi \rangle = \langle \partial\delta, \phi \rangle = -\langle \delta, \partial\phi \rangle = -\phi'(0)$. More generally, putting $H_a(x) = H(x-a)$ and $\delta_a(x) = \delta(x-a)$, we have

$$\langle \partial H_a, \phi \rangle = \langle \delta_a, \phi \rangle = \phi(a),$$

and

$$\langle \partial^k \delta_a, \phi \rangle = (-1)^k \phi^{(k)}(a).$$

Let us now consider an example of a product of two functions. Let $f \in C^\infty(\mathbb{R})$ and let H be the Heaviside step function as above, and consider $Hf(x) := H(x)f(x)$. We compute the distributional derivative of Hf as follows:

$$\begin{aligned} \langle \partial(Hf), \phi \rangle &= -\langle Hf, \partial\phi \rangle \\ &= -\int_{\mathbb{R}} H(x)f(x)\phi'(x) \\ &= -\int_0^\infty f(x)\phi'(x)\,\mathrm{d}x \\ &= f(0)\phi(0) + \int_0^\infty f'(x)\phi(x)\,\mathrm{d}x \\ &= f(0)\phi(0) + \int_{\mathbb{R}} f'(x)H(x)\phi(x)\,\mathrm{d}x \\ &= \langle f(0)\delta, \phi \rangle + \langle f'H, \phi \rangle, \end{aligned}$$

so $\partial(Hf) = f(0)\delta + f'H$. Note that we have *not* found that $\partial(Hf) = (\partial H)f + f'H$; that is, the product rule does *not* hold for distributional derivatives. Indeed, by induction we can compute that

$$\partial^k(Hf) = \sum_{j=0}^{k-1} \partial^j f(0)\partial^{k-j-1}\delta + H\partial^k f.$$

Try as we might, the product of two distributions doesn't make sense. However, we *can* make sense of the product of a smooth function and a distribution. Given two smooth functions $f \in C_c^\infty(X)$ and $u \in C_c^\infty(X)$, and considering u as a distribution, we can see that

$$\langle fu, \phi \rangle = \int_X fu \cdot \phi \, \mathrm{d}x = \int_X u \cdot f\phi \, \mathrm{d}x = \langle u, f\phi \rangle.$$

So, given $f \in C_c^\infty(X)$ and $u \in \mathcal{D}'(X)$, it is natural to *define*

$$\langle fu, \phi \rangle := \langle u, f\phi \rangle.$$

One must check that this does indeed define a distribution. Let ϕ_j be a sequence of functions in $C_c^\infty(X)$ such that $\phi_j \to \phi$. Then, as $f \in C_c^\infty(X)$, we have that $f\phi_j \to f\phi$ in $C_c^\infty(X)$, so that

$$\langle fu, \phi_j \rangle = \langle u, f\phi_j \rangle \to \langle u, f\phi \rangle = \langle fu, \phi \rangle,$$

so fu is a distribution. Indeed, this definition makes sense even if $f \in C^\infty(X)$, since then $f\phi \in C_c^\infty(X)$. However, there is no obvious way of reducing the regularity of f. Even so, counterintuitive results can appear: given $f \in C_c^\infty(\mathbb{R})$, let us consider the product $f\delta$. We see that

$$\langle f\delta, \phi \rangle = \langle \delta, f\phi \rangle = f(0)\phi(0) = f(0)\langle \delta, \phi \rangle.$$

So $f\delta$ is the same as $f(0)\delta$!

TO FINISH

3.4 Distributions of Compact Support, Tensor Products and Convolutions

3.5 Fourier Transform of Tempered Distributions*

3.6 Sobolev Spaces*

3.7 Fundamental Solutions

4 Hilbert Transform*

References

[Duo] Javier Duoandikoetxea. *Fourier analysis*. Vol. 29. Graduate Studies in Mathematics. Translated and revised from the 1995 Spanish original by David Cruz-Uribe. Providence, RI: American Mathematical Society, 2001, pp. xviii+222. ISBN: 0-8218-2172-5.

[Fol] Gerald B. Folland. *Fourier analysis and its applications*. The Wadsworth & Brooks/Cole Mathematics Series. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1992, pp. x+433. ISBN: 0-534-17094-3.

[Fri&Jos] F. G. Friedlander. *Introduction to the theory of distributions*. Second edition. With additional material by M. Joshi. Cambridge: Cambridge University Press, 1998, pp. x+175. ISBN: 0-521-64015-6; 0-521-64971-4.

[GraCl] Loukas Grafakos. *Classical Fourier analysis*. Second edition. Vol. 249. Graduate Texts in Mathematics. New York: Springer, 2008, pp. xvi+489. ISBN: 978-0-387-09431-1.

[GraMo] Loukas Grafakos. *Modern Fourier analysis*. Second edition. Vol. 250. Graduate Texts in Mathematics. New York: Springer, 2009, pp. xvi+504. ISBN: 978-0-387-09433-5.

[Rud] Walter Rudin. *Functional analysis*. Second edition. International Series in Pure and Applied Mathematics. New York: McGraw-Hill Inc., 1991, pp. xviii+424. ISBN: 0-07-054236-8.

[SteHA] Elias M. Stein. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. Vol. 43. Princeton Mathematical Series. With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III. Princeton, NJ: Princeton University Press, 1993, pp. xiv+695. ISBN: 0-691-03216-5.

[Ste&Sha] Elias M. Stein and Rami Shakarchi. *Fourier analysis*. Vol. 1. Princeton Lectures in Analysis. An introduction. Princeton, NJ: Princeton University Press, 2003, pp. xvi+311. ISBN: 0-691-11384-X.

[SteSI] Elias M. Stein. *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30. Princeton, N.J.: Princeton University Press, 1970, pp. xiv+290.

[Yos] Kōsaku Yosida. *Functional analysis*. Classics in Mathematics. Reprint of the sixth (1980) edition. Berlin: Springer-Verlag, 1995, pp. xii+501. ISBN: 3-540-58654-7.