Math 725

Lecture notes

Spring 2000

This document is available in pdf format at the following website:
`www.math.wisc.edu/~ angenent/725`

The TEX files are available on request

Contents

A very short summary of the theory of Distributions

Textbooks to look at: Both [5, 8] offer a chapter introducing you to the theory of distributions, but are written with the assumption that you know something about locally convex topological vector spaces. The shorter text [3] is entirely devoted to distributions and, like these notes and the material presented in class, it is written on a more elementary level.

1. Test functions

Let Ω be an open subset of $\mathbb{R}^d$. A *test function on* Ω is a function $\varphi : \Omega \to \mathbb{R}$ for which

1. $\operatorname{supp}\varphi$ is compact, and
2. φ has derivatives of all orders

The space of test functions on Ω is denoted by $\mathcal{D}(\Omega)$.

Exercise 1. If $\varphi, \psi \in \mathcal{D}(\Omega)$ then $a\varphi + b\psi \in \mathcal{D}(\Omega)$ for any $a, b \in \mathbb{R}$.

1.1. Convergence of test functions.

A sequence of test functions $\varphi_n \in \mathcal{D}(\Omega)$ converges in $\mathcal{D}(\Omega)$ to φ if

1. there is a compact $K \subset \Omega$ with $\operatorname{supp}\varphi_n \subset K$ for all n,
2. all derivatives $\partial^\alpha \varphi_n$ converge uniformly to $\partial^\alpha \varphi$

Here I've used the "multi-index notation" for partial derivatives, where

$$\partial^\alpha \varphi(x) = \frac{\partial^{a_1+\ldots+a_d}\,\varphi}{(\partial x_1)^{a_1}\ldots(\partial x_d)^{a_d}}$$

and where the so-called "multi-index" $\alpha = (a_1, \ldots, a_d)$ is a d-tuple of nonnegative integers. One defines $|\alpha| = a_1 + \ldots + a_d$.

2. Distributions

A distribution is a linear map $T : \mathcal{D}(\Omega) \to \mathbb{R}$ which is continuous in the sense that

$$\lim_{n\to\infty} T(\varphi_n) = T(\varphi) \text{ for any sequence } \varphi_n \overset{\mathcal{D}(\Omega)}{\longrightarrow} \varphi.$$

The space of distributions on Ω is denoted by $\mathcal{D}'(\Omega)$.

3. Examples

3.1. Locally integrable functions.

Let $f \in L^1_{\text{loc}}(\Omega)$ be given. Then we define $T_f \in \mathcal{D}'(\Omega)$ by setting

$$T_f(\varphi) = \int_\Omega f(x)\varphi(x)\, dx.$$

Exercise 2. Show that T_f is a distribution, in particular verify the continuity condition $T_f(\varphi_n) \to T_f(\varphi)$.

Lemma 1. *If f, g are locally integrable functions which are the same almost everywhere, then T_f and T_g are the same distribution.*

Conversely, if $T_f = T_g$ then $f = g$ a.e.

Proof. For $h = f - g$ we have $T_h = T_f - T_g = 0$, which means that

$$\int_\Omega h(x)\varphi(x)\, dx = 0 \text{ for all } \varphi \in \mathcal{D}(\Omega).$$

This implies $h(x) = 0$ a.e. (I postpone the proof of this statement to section 7.) □

3.2. Dirac's delta function.

The distribution defined by

$$\langle \delta, \varphi \rangle = \varphi(0), \qquad \forall \varphi \in \mathcal{D}(\mathbb{R}),$$

is called Dirac's delta function.

Exercise 3.
(i) Verify that δ is a distribution.
(ii) Show that $\delta \neq T_f$ for any locally integrable function f on $\mathbb{R}$.

Generalizations of Dirac's delta function are Borel measures on Ω. If μ is a Radon measure on Ω, then

$$\langle \mu, \phi \rangle = \int_\Omega \varphi(x) d\mu(x)$$

again defines a distribution.

For example, if $\Gamma \subset \mathbb{R}^2$ is a curve, and ds is arc length along the curve, then

$$\langle T, \varphi \rangle = \int_\Gamma \varphi\, ds$$

defines a distribution on $\mathbb{R}^2$. In this case the associated measure μ is given by $\mu(E) =$ length of the portion of the curve Γ contained in the set E

3.3. Principal Value of $1/x$.

The function $f(x) = 1/x$ is measurable but not locally integrable near $x = 0$. Nevertheless, one can define a distribution by

$$\langle T, \varphi \rangle = \lim_{\varepsilon \searrow 0} \int_{|x| \geq \varepsilon} \frac{\varphi(x)}{x}\, dx.$$

Exercise 4. Prove that the limit exists, and show that it defines a distribution in $\mathcal{D}'(\mathbb{R})$.

The distribution thus defined is called the Cauchy Principal Value of $1/x$, and is denoted by $\text{P.V.}\dfrac{1}{x}$.

Exercise 5. The limit

$$\langle S, \varphi \rangle = \lim_{\varepsilon \searrow 0} \int_{-\infty}^{-\varepsilon} + \int_{2\varepsilon}^{\infty} \frac{\varphi(x)}{x}\, dx$$

also defines a distribution.

Compute $S - T$, where T is as above.

4. Derivatives of distributions

If $T \in \mathcal{D}'(\Omega)$ is a distribution, then we define its partial derivative with respect to x_i to be the distribution $D_i T$, specified by

$$\langle D_i T, \varphi \rangle = -\langle T, \frac{\partial \varphi}{\partial x_i} \rangle.$$

Exercise 6. Check that $D_i T$ does indeed define a distribution.

4.1. Consistency of the definition.

If $T = T_f$, and f is a continuously differentiable function then we now have two defitions of the partial derivatives of f. Integration by parts shows that

$$T_{\partial f/\partial x_i} = D_i T_f,$$

indeed, for any $\varphi \in \mathcal{D}(\Omega)$ we have

$$\begin{aligned}
\langle T_{\partial f/\partial x_i}, \varphi \rangle &= \int \frac{\partial f}{\partial x_i} \varphi \, dx \\
&= \int \left\{ \frac{\partial f \varphi}{\partial x_i} - f \frac{\partial \varphi}{\partial x_i} \right\} dx \\
&= -\int f \frac{\partial \varphi}{\partial x_i}\, dx \\
&= -\langle T_f, \frac{\partial \varphi}{\partial x_i} \rangle \\
&= \langle D_i T_f, \varphi \rangle.
\end{aligned}$$

Here we have used the fact that if g is a continuously differentiable function with compact support in Ω, then

$$\int_\Omega \frac{\partial g}{\partial x_i}\, dx = 0.$$

Since the new and old definitions for derivative coincide for continuously differentiable functions we will use any of the usual notations for derivatives, i.e. $\partial_i f = D_i f = f_{x_i} = \frac{\partial f}{\partial x_i}$.

Once one has defined the derivative one can define higher derivatives by induction, e.g. $D_i D_j T$ is $D_i(D_j T)$.

Exercise 7. Show that for distributions $D_i D_j T = D_j D_i T$, without any further restrictions on $T \in \mathcal{D}'(\Omega)$.

4.2. Derivatives of the delta function.

Applying the definition one finds that the derivative of the Dirac δ is given by

$$\langle D\delta, \varphi \rangle = -\varphi'(0).$$

The n^{th} derivative is given by

$$\langle D^n \delta, \varphi \rangle = (-1)^n \varphi^{(n)}(0).$$

4.3. Other derivative examples.

Exercise 8. Show that $\delta = DT$ where $T = T_{\chi_{[0,\infty)}}$.
Compute the second derivative in $\mathcal{D}'(\mathbb{R})$ of $f(x) = |x|$.

Exercise 9. Show, by integrating by parts, that

$$\text{P.V.}\frac{1}{x} = DT$$

where $T = T_{\ln|x|}$.

Exercise 10. Let f be the measurable function defined on $\mathbb{R}^d$ by $f(x) = |x|^{-a}$, in which a is a positive constant, and where $|x| = \sqrt{(x_1^2 + \ldots + x_d^2)}$.
For which $a > 0$ is f locally integrable?
For $x \neq 0$ one has

$$\partial_i f(x) = -a \, \frac{x_i}{|x|^{a+2}}.$$

For which values of a does the righthandside define a locally integrable function, and does the equation also hold in the sense of distributions?

Exercise 11. Let $a = a_0 < a_1 < \ldots < a_n = b$ be given real numbers. Given k functions $f_i \in C^2([a_{i-1}, a_i])$ $(i = 1, \ldots, k)$ defined on adjacent intervals we consider the piecewise continuous function $f : (a, b) \to \mathbb{R}$

$$f(x) = f_i(x) \text{ if } x \in (a_{i-1}, a_i).$$

At $x = a_i$ we define $f(x) = 0$.
(i) Compute $Df \in \mathcal{D}'(\Omega)$, where $\Omega = (a, b)$.
(ii) Compute $D^2 f \in \mathcal{D}'(\Omega)$.
(iii) When is Df a locally integrable function?
(iv) When is $D^2 f$ a locally integrable function?

Exercise 12. Let $E \subset \mathbb{R}^2$ be a bounded subset whose boundary is a differentiable curve (e.g. E is a disc). Then the characteristic function $\chi_E(x)$ of the set E is a locally integrable function, and thus defines a distribution. Compute $D_i \chi_E$ for $i = 1, 2$.

Exercise 13. For any constant $K \in \mathbb{R}$ we let $f(x,y) = K - x^2 - y^2$ for $x^2 + y^2 < 1$ and $f(x,y) = 0$ elsewhere in $\mathbb{R}^2$.
(i) Compute $D_x f$ and $D_y f$.
(ii) For which value(s) of K are $D_x f$ and $D_y f$ locally integrable functions?
(iii) Assume K is such that $D_x f, D_y f \in L^1_{\mathrm{loc}}(\mathbb{R}^2)$, and compute $\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$.

Exercise 14. Let $E \subset \mathbb{R}^d$ be a bounded subset whose boundary ∂E is smooth. Let $\tilde{f} : E \to \mathbb{R}$ be a C^2 function which vanishes on ∂E, i.e. $\tilde{f}(x) = 0$ for all $x \in \partial E$. Consider the function

$$f(x) = \begin{cases} \tilde{f}(x) & \text{if } x \in E \\ 0 & \text{elsewhere} \end{cases}$$

Compute Δf.

5. Convergence of distributions

A sequence of distributions $T_n \in \mathcal{D}'(\Omega)$ converges *in the sense of distributions* to $T \in \mathcal{D}'(\Omega)$, if for every test function $\varphi \in \mathcal{D}(\Omega)$ one has

$$\lim_{n\to\infty} \langle T_n, \varphi \rangle = \langle T, \varphi \rangle.$$

Notation: $T_n \overset{\mathcal{D}'(\Omega)}{\longrightarrow} T$, or "$T_n \to T$ in $\mathcal{D}'(\Omega)$."

Exercise 15. Show that if $T_n \overset{\mathcal{D}'(\Omega)}{\longrightarrow} T$ then $D_i T_n \overset{\mathcal{D}'(\Omega)}{\longrightarrow} D_i T$.

A typical example is this: Let $\Omega = \mathbb{R}$ and consider$f_n(x) = \frac{1}{n} \sin nx$. Then f_n converges uniformly to zero, and hence $T_{f_n} \overset{\mathcal{D}'(\Omega)}{\longrightarrow} 0$.

By the previous exercise the derivative $\cos nx = D(n^{-1} \sin nx)$ also converges to zero in $\mathcal{D}'(\Omega)$! This is an instance of the "Riemann-Lebesgue Lemma" which states

$$\lim_{n\to\infty} \int_{\mathbb{R}} \varphi(x) \cos nx \, dx = 0$$

for any $\varphi \in L^1(\mathbb{R})$. Here we have proved this for the case that φ is a test function, $\varphi \in \mathcal{D}(\mathbb{R})$.

Exercise 16. Show that $n^{1999} \sin nx \to 0$ in $\mathcal{D}'(\mathbb{R})$ as $n \to \infty$.

The following problem has perhaps a surprising answer:

Exercise 17. If $f_n(x) = \cos nx$ then $f_n \to 0$ in $\mathcal{D}'(\mathbb{R})$. Compute the limit in $\mathcal{D}'(\mathbb{R})$ of $g_n(x) = (f_n(x))^2 = \cos^2 nx$. *(Hint: use a double angle formula)*

6. Multiplication of distributions with smooth functions

In general one cannot multiply distributions with each other in the same way that old fashioned functions can be multiplied. The best one can do for distributions is this: If $T \in \mathcal{D}'(\Omega)$ and if $g \in C^\infty(\Omega)$ then the product $g \cdot T$ is defined by

$$\langle gT, \varphi \rangle = \langle T, g\varphi \rangle.$$

The crucial remark here is that $g\varphi$ is again a test function for any test function φ.

Exercise 18. Show that $gT_f = T_{gf}$ for any locally integrable f.

Exercise 19. Prove the product rule, i.e. show that if $T \in \mathcal{D}'(\Omega)$ and $f \in C^\infty(\Omega)$ then one has

$$D_i(fT) = fD_iT + (D_if)T.$$

The following example shows that, in order to define the product of a function $f(x)$ and a distribution ($\delta^{(n)}(x)$ in this case) one may need derivatives of $f(x)$ of arbitrary high order. This indicates why one cannot expect to give a good definition of the product of a function f with *any distributution* $T \in \mathcal{D}(\mathbb{R})$ if the function only has a finite number of derivatives.

Exercise 20.

(i) Prove

$$x\delta(x) = 0$$
$$x\delta'(x) = -\delta(x)$$
$$x\delta''(x) = -2\delta'(x)$$

(ii) Show that for any $f \in C^\infty(\mathbb{R})$ one has

$$f(x)\delta(x) = f(0)\delta(x),$$
$$f(x)\delta'(x) = f(0)\delta'(x) - f'(0)\delta(x),$$
$$f(x)\delta''(x) = f(0)\delta''(x) - 2f'(0)\delta'(x) + f''(0)\delta(x)$$

(iii) Show that for any $f \in C^\infty(\mathbb{R})$ one has and any $n \in \mathbb{N}$ one has

$$f(x) \cdot \delta^{(n)}(x) = \sum_{k=0}^{n} \binom{n}{k} (-1)^k f^{(n-k)}(0)\delta^{(k)}(x).$$

where both sides are interpreted as distributions in $\mathcal{D}'(\mathbb{R})$.

7. Appendix: Examples of test functions

The function

$$\Phi(x) = \begin{cases} e^{-1/x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

is a C^∞ function on $\mathbb{R}$. This function is monotone nondecreasing, and satisfies

$$\lim_{x \to \infty} \Phi(x) = 1.$$

For any positive λ and any interval (a, b) the function

$$\Phi_{\lambda,a,b}(x) = \Phi(\lambda(b - x)(x - a))$$

is strictly positive on the interval (a, b) and zero elsewhere. It is the composition of C^∞ functions and hence again C^∞.

As $\lambda \nearrow \infty$ the functions $\Phi_{\lambda,a,b}(x)$ converge monotonically to the characteristic function $\chi_{(a,b)}$ of the interval (a, b).

Thus if $h(x)$ is a locally integrable function on $\mathbb{R}$ for which $\int_{\mathbb{R}} h(x)\varphi(x)dx = 0$ for all $\varphi \in \mathcal{D}(\mathbb{R})$, then this also holds for all $\varphi = \Phi_{\lambda,a,b}$, and by taking the limit $\lambda \nearrow \infty$ the dominated convergence theorem implies that $\int_{(a,b)} h(x)dx = 0$ for every interval (a, b). This implies that $h(x) = 0$ a.e.

For functions of several variables one can use the same arguments and thus prove:

Theorem 2. *If $h \in L^1_{\mathrm{loc}}(\Omega)$ satisfies $\int h(x)\varphi(x)\ dx = 0$ for all test functions $\varphi \in \mathcal{D}(\Omega)$ then $h(x) = 0$ a.e.*

Proof. For every "rectangle" $(a_1, b_1) \times \ldots \times (a_d, b_d)$ one considers the function

$$\Phi_\lambda(x_1, \ldots, x_d) = \prod_{i=1}^{d} \Phi_{\lambda, a_i, b_i}(x_i).$$

As in the one dimensional case these $\Phi_\lambda(x)$ converge monotonically to the characteristic function of the rectangle $\mathcal{R} = (a_1, b_1) \times \ldots \times (a_d, b_d)$.

The hypothesis that $\int h\varphi dx = 0$ for every $\varphi \in \mathcal{D}(\Omega)$ then implies $\int h\Phi_\lambda dx = 0$ for any $\lambda > 0$, and any rectangle $\mathcal{R}$ which is contained in Ω. Letting $\lambda \nearrow \infty$ again we conclude that $\int_{\mathcal{R}} h(x)dx = 0$ for every rectangle $\mathcal{R} \subset \Omega$. The theory of Lebesgue integration then implies that $h(x) = 0$ a.e. in Ω. □

8. Appendix: A Motivating example

In the theory of conformal mappings one encounters the following so-called "Dirichlet-problem": Given a domain $\Omega \subset \mathbb{R}^2$ with a smooth boundary $\partial\Omega$, and a function $g : \partial\Omega \to \mathbb{R}$ defined one this boundary, find a function $f : \Omega \to \mathbb{R}$ which satisfies Laplace's equation

$$\Delta f \stackrel{\text{def}}{=} \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$$

in the domain Ω, and which equals g on $\partial\Omega$.

Dirichlet observed that if there is a solution f_* to this problem, then it minimizes the Dirichlet integral

$$D(f) \stackrel{\text{def}}{=} \frac{1}{2}\int_\Omega |\nabla f(x)|^2\ dx$$

among all functions f with $f(x) = g(x)$ for all $x \in \partial\Omega$. Conversely, if a given f_* minimizes $D(f)$, then it must be a solution to the Dirichlet problem.

Here we will prove this converse, assuming that f_* is only a C^1 function.

Theorem 3. *Let X be the set of functions $f : \bar{\Omega} \to \mathbb{R}$ which are continuous on $\bar{\Omega}$ and which have continuous first derivatives in Ω. Assume that $f_* \in X$ minimizes $D(f)$ among all $f \in X$ with $f = g$ on $\partial\Omega$. Then f_* satisfies Laplace's equation in the sense of distributions.*

Proof. Let $\varphi \in \mathcal{D}(\Omega)$ be a test function. Then $f_* + t\varphi \in X$ for all $t \in \mathbb{R}$ and so $D(f_* + t\varphi) \geq D(f_*)$ for all $t \in \mathbb{R}$. Moreover $D(f_* + t\varphi)$ attains its minimum value at $t = 0$. Now exand $D(f_* + t\varphi)$:

$$\begin{aligned} D(f_* + t\varphi) &= \frac{1}{2}\int |\nabla(f_* + t\varphi)|^2\ dx \\ &= \frac{1}{2}\int \{|\nabla f_*|^2 + 2\nabla f_* \cdot \nabla\varphi + |\nabla\varphi|^2\}\ dx \end{aligned}$$

and compute the derivative of this expression at $t = 0$

$$\begin{aligned}\left.\frac{dD(f_* + t\varphi)}{dt}\right|_{t=0} &= \int \nabla f_* \cdot \nabla\varphi \, dx \\ &= \int \{\partial_x f_* \partial_x \varphi + \partial_y f_* \partial_y \varphi\} \, dx \\ &= \langle \partial_x f_*, \partial_x \varphi \rangle + \langle \partial_y f_*, \partial_y \varphi \rangle \\ &= \langle -\Delta f_*, \varphi \rangle\end{aligned}$$

Since this derivative must vanish we see that $\Delta f_* = 0$ in the sense of distributions. □

If you look carefully at the proof then you see that the same argument shows that any f_* which minimizes $D(f)$, and whose second derivatives are continuous must be a solution to Laplace's equation in the ordinary, non-distribution, sense. So what did we gain by using the theory of distributions here?

If you try to construct a solution to Dirichlet's problem by showing that there exists a function f_* which minimizes $D(f)$ then it is not clear a priori that such a minimizer will be C^2 rather than C^1. It may be easier to find a C^1 minimizer than a C^2 minimizer. The theory of distributions allows us to state that even C^1 minimizers satisfy Laplace's equation, at least in a generalized sense.

Exercise 21. Let $f_i \in C(\bar{\Omega})$ be a sequence of solutions (in the sense of distributions) of Laplace's equation, and assume that the f_i converge uniformly to some function f_∞. Prove that f_∞ is again a solution of Laplace's equation.

Banach Spaces

Norms and Seminorms, Equivalent norms, Finite dimensional examples, An infinite dimensional Banach space, Bounded linear operators and the dual space, Linear subspaces, The unit ball, Series in Banach spaces, Sums and quotients of Banach spaces.

Text books to look at: In the library there are many books called "Functional Analysis," and almost all of them present the theory of Banach spaces in varying degrees of detail. The material in this section is usually found in a first chapter of such a book. Books you could look at are [4, 5, 6, 8].

Rudin's [4] is a good choice to read next to these notes, in particular because I'll follow parts of this book later on in the course. Rudin's other book [5] gives a more comprehensive account of Functional Analysis, but it takes the theory of Locally Convex Topological Vector Spaces as its starting point, which may make for difficult first reading, and is a level of generality we won't pursue in this course.

9. Norms and Seminorms

A *seminorm* on a vector space X is a nonnegative function $p : X \to \mathbb{R}$ which is *homogeneous*,

$$p(\lambda x) = |\lambda| p(x) \text{for all } x \in X, \text{ and all } \lambda \in \mathbb{R},$$

and subadditive,

$$p(x + y) \leq p(x) + p(y) \text{ for all } x, y \in X.$$

A seminorm $p : X \to \mathbb{R}$ is a *norm* if it satisfies

$$p(x) = 0 \Leftrightarrow x = 0.$$

Norms are usually denoted by $\|x\|_X$ with a subscript to indicate which norm, if confusion is possible.

A norm defines a metric (distance function) by

$$d_X(x, y) = \|x - y\|_X.$$

A *Banach space* is a normed vector space $(X, \|\cdot\|)$ which is complete for the metric d_X. Recall that completeness means that every Cauchy sequence $\{x_i \in X\}_{i \in \mathbb{N}}$ must have a limit.

10. Equivalent norms

Two norms $\|\dots\|_1$ and $\|\dots\|_2$ on the same vector space X are called *equivalent* if there exist constants $c, C > 0$ such that

$$c\|x\|_1 \leq \|x\|_2 \leq C\|x\|_1 \text{ for all } x \in X.$$

If $\|\dots\|_1$ and $\|\dots\|_2$ are equivalent norms then sequences $\{x_i : i \in \mathbb{N}\}$ converge in one norm if and only if the converge in the other. Put differently, we have two distance functions $d_i(x,y) = \|x-y\|_i$ on X, and the identity map $\mathrm{id}_X : X \to X$ is a homeomorphism from (X, d_1) to (X, d_2).

The same definition may be applied to seminorms instead of norms.

Exercise 22. Let X be a vector space and $\{p_s : s \in S\}$ a collection of seminorms on X. Show that, if

$$q(x) = \sup_{s \in S} p_s(x)$$

is finite for all $x \in X$, then q is again a seminorm; i.e. "the sup of seminorms is again a seminorm."

(No assumption on the size of S is intended, S could be finite , countable, or uncountable.)

11. Finite dimensional examples

If X is a finite dimensional vectorspace then we may choose a basis and identify X with $\mathbb{R}^N$. The quantity

$$p(x_1, \dots, x_N) = |x_1|$$

defines a seminorm. The quantities

$$p_\infty(x) = \max\{|x_i| : i = 1, 2, \dots, N\} \qquad \text{(maximum norm)}$$
$$p_1(x) = \textstyle\sum_{i=1}^N |x_i| \qquad \text{(sum norm)}$$
$$p_2(x) = |x| = \sqrt{x_1^2 + \dots + x_N^2} \qquad \text{(Euclidean length)}$$

all define norms on $\mathbb{R}^N$.

Exercise 23. Show that the norms p_1, p_2, and p_∞ are pairwise equivalent and find the constants "c, C."

Exercise 24. (Constructing new seminorms from old ones.)
If $p_1, p_2, \dots, p_N : X \to \mathbb{R}$ are seminorms on a vector space X, then the quantities

$$q(x) = \textstyle\sum_i p_i(x),$$
$$r(x) = \max_i p_i(x),$$
$$s(x) = \sqrt{p_1(x)^2 + \dots + p_N(x)^2}$$

are also seminorms on X.

Show this and also show that these seminorms are equivalent.

Exercise 25. A function $f : X \to \mathbb{R}$ is called **convex** if it satisfies

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$

for all $x, y \in X$ and $0 \leq t \leq 1$.

Show that any seminorm is convex, and conversely that any convex function $f : X \to \mathbb{R}$ which is homogeneous ($f(\lambda x) = |\lambda| f(x)$ for all $\lambda \in \mathbb{R}$) is a seminorm.

The following theorem shows that all norms define the same topology on $\mathbb{R}^N$.

Theorem 4. *Every norm $p(x)$ on $\mathbb{R}^N$ is equivalent to the Euclidean norm $p_2(x)$.*

Proof. We will denote the Euclidean norm by $|x|$, as is more customary.

If $e_1, \ldots, e_N$ is the standard basis for $\mathbb{R}^N$ then any vector x is of the form $x = \sum_i x_i e_i$, and one has

$$\begin{aligned} p(x) &= p(x_1 e_1 + \ldots + x_N e_N) \\ &\leq |x_1| p(e_1) + \ldots + |x_N| p(e_N) \\ &\leq C\sqrt{x_1^2 + \ldots + x_N^2}, \qquad \text{(Cauchy-Schwarz } \leq) \\ &= C|x|. \end{aligned}$$

where $C = \sqrt{(p(e_1)^2 + \ldots + p(e_N)^2)}$.

This implies that the norm p is a continuous function on $\mathbb{R}^N$, since

$$|p(x) - p(y)| \leq p(x - y) \leq C|x - y|.$$

Thus p is a continuous function which is strictly positive on the unit sphere $S = \{x \in \mathbb{R}^N : |x| = 1\}$. This sphere is compact, and hence $p|_S$ is bounded from below by a positive constant, i.e.

$$\inf_{x \in S} p(x) = c > 0.$$

Consequently we have

$$p(x) = |x| p\left(\frac{x}{|x|}\right) \geq c|x|$$

for all $x \in \mathbb{R}^N$. The norms $p(x)$ and $|x|$ are therefore equivalent. □

Exercise 26.
(i) We just used the following inequality: $|p(x) - p(y)| \leq p(x - y)$. Derive this from the axioms of a seminorm.
(ii) How does $|p(x) - p(y)| \leq C|x - y|$ imply that p is continuous?

12. An infinite dimensional Banach space

Let K be a compact metric space, e.g. K could be a compact subset of $\mathbb{R}^d$, or even $K = [0, 1]$. Then $C(K)$ is by definition the set of continuous functions on K. This is a vector space, and the quantity

$$\|f\| = \sup_{x \in K} |f(x)|$$

defines a norm. A sequence of functions f_i converges to some $f \in C(K)$ exactly if the f_i converge uniformly on K to f.

Exercise 27. Verify that $\|\ldots\|$ is indeed a norm, and show that $C(K)$ is a Banach space with this norm (i.e. verify that $C(K)$ is complete.)

13. Bounded linear operators and the dual space

Let X and Y be Banach spaces, and let $T : X \to Y$ be a linear mapping. Then T is called *bounded* if

$$\|Tx\|_Y \leq C\|x\|_X$$

for some constant $C < \infty$ which does not depend on $x \in X$.

The smallest constant C one can take is called the *operator norm* of T. Equivalent expressions for the operator norm of T are

$$\|T\| \stackrel{\text{def}}{=} \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X \leq 1} \|Tx\|_Y = \sup_{\|x\|_X = 1} \|Tx\|_Y.$$

Lemma 5. *A linear map $T : X \to Y$ of normed vector spaces is continuous if and only if it is bounded.*

Proof. Suppose T is bounded. Let $x \in X$ and $\varepsilon > 0$ be given. Then choose $\delta = \varepsilon / \|T\|$ and observe that $\|x' - x\|_X < \delta$ implies

$$\|Tx' - Tx\|_Y = \|T(x' - x)\|_Y \leq \|T\| \cdot \|x' - x\|_X < \varepsilon.$$

Hence T is continuous at $x \in X$.

Conversely, suppose T is continuous. Since T is linear one has $T(0) = 0$, and hence there exists a $\delta > 0$ such that $\|x\|_X < \delta$ implies $\|Tx\|_Y < 1$. For arbitrary $x \in X$ we then have

$$\|Tx\| = \left\| \frac{\|x\|_X}{\delta} T\left(\delta \frac{x}{\|x\|_X}\right) \right\| \leq \frac{\|x\|_X}{\delta},$$

so that T is bounded with $\|T\| \leq \delta^{-1}$. □

The space of bounded operators from X to Y is denoted by $\mathcal{L}(X, Y)$. With the operator norm $\mathcal{L}(X, Y)$ is a normed vector space.

Different notation and terminology is used in the special case which you get by choosing $Y = \mathbb{R}$ (the real numbers with "norm" given by $\|x\| = |x|$ is a Banach space!). A linear map $T : X \to \mathbb{R}$ is called a "linear functional," its "operator norm" is defined in the same way,

$$\|T\| = \sup_{\|x\| \leq 1} |T(x)|,$$

and the space of bounded linear functionals $T : X \to \mathbb{R}$ is called the *dual space* of X. It is denoted by $X^* = \mathcal{L}(X, \mathbb{R})$.

Exercise 28. Verify that the operator norm is indeed a norm.

Finally, one has the following important observation.

Theorem 6. *If X is a normed vector space and Y is a Banach space then $\mathcal{L}(X, Y)$ with the operator norm is a Banach space.*

In particular, the dual X^* of any normed vector space X is a Banach space.

You could provide a proof yourself, or use the absolutely convergent series approach, which will be discussed shortly, to prove completeness.

14. Linear subspaces

A subset $L \subset X$ is a linear subspace if $ax+by \in L$ for all $x, y \in L$ and $a, b \in \mathbb{R}$.

If X is finite dimensional, i.e. if $X = \mathbb{R}^N$ with some norm, then all linear subspaces of X are closed subsets. In infinite dimensional spaces this is not always true as the following example shows.

Let $X = C(K)$ with $K = [0,1]$, and let L be the space consisting of all polynomials $f(x) = a_m x^m + \ldots + a_1 x + a_0$. Then clearly $L \neq X$, but the Weierstrass approximation theorem states that the closure of L is X, i.e. every continuous function can be approximated uniformly by polynomials.

15. The unit ball

The set $B_X = \{x \in X : \|x\| \leq 1\}$ has the following four properties:

1. It is convex, i.e. the linesegment connecting any two points $x, y \in B_X$ is again in B_X,
2. it is symmetric: $x \in B_X$ if and only if $-x \in B_X$,
3. it is absorbing, meaning that every $x \in X$ is contained in some homothetic copy $tB_X \stackrel{\text{def}}{=} \{tx : x \in B_X\}$ with $t > 0$ of the unit ball.
4. if $tx \in B_X$ for all $t > 0$ then $x = 0$

Conversely, if $B \subset X$ is a set satisfying these four properties then

$$p_B(x) = \inf\{t > 0 : x \in tB\}$$

defines a norm on X.

Exercise 29. Verify these statements.

Exercise 30. Draw the unit balls in $\mathbb{R}^3$ for the norms p_1, p_2 and p_∞

Lemma 7. *The unit ball B_X is compact if and only if X is finite dimensional.*

Proof. If X is finite dimensional then $X = \mathbb{R}^N$ for some N, and B_X is a closed and bounded subset of X. The Bolzano-Weierstrass-Heine-Borel theorem implies that B_X is compact.

Suppose X is not finite dimensional. Then we will construct a sequence of vectors $x_i \in X$ with $\|x_i\| = 1$ and $\|x_i - x_j\| = 1$ for all $i \neq j$. Such a sequence is bounded but has no convergent subsequence, so that B_X is not compact.

To construct the x_i we choose x_1 with $\|x_1\| = 1$, but arbitrarily otherwise. Assuming the first n vectors $x_1, \ldots, x_n$ have been constructed we let L be the linear subspace of X spanned by $\{x_1, \ldots, x_n\}$. Since X is infinite dimensional L is a proper subset of X and hence a vector $y \in X \setminus L$ exists.

Let $z \in L$ be a point which minimizes the distance $\|y - z\|$. To see that such a z must exist we consider the function $f : L \to \mathbb{R}$ given by $f(z) = \|z - y\|$. Let $R = \|y\|$. Then, since L is finite dimensional, the set $K = \{z \in L : \|z\| \leq 3R\}$ is compact, and f attains a minimum on K, say at some $z_* \in K$. This minimum value cannot exceed $f(0) = \|y\| = R$. On the complement of K, i.e. on $L \setminus K$ one has

$$f(z) = \|z - y\| \geq \|z\| - \|y\| \geq 3R - R = 2R > f(z_*).$$

So z_* is a nearest point to y in L.

We now define

$$x_{n+1} = \frac{y - z_*}{\|y - z_*\|}.$$

Then x_{n+1} is clearly a unit vector, and by construction its distance to L also equals 1. Hence $\|x_{n+1} - x_i\| = 1$ for $1 \le i \le n$. □

16. Series in Banach spaces

A series $\sum_{i=1}^{\infty} x_i$ whose terms lie in a normed vector space X converges if the partial sums $s_N = \sum_1^N x_i$ converge. The limit of the partial sums is the sum of the series,

$$\textstyle\sum_{i=1}^{\infty} x_i = \lim_{N\to\infty} \sum_1^N x_i.$$

A series $\sum_i x_i$ is called *absolutely convergent* if the series $\sum_i \|x_i\|$ converges (in $\mathbb{R}$).

Theorem 8. *A normed vector space X is complete if and only if every absolutely convergent series in X converges.*

Proof. If X is complete (X is a Banach space), then the sequence of partial sums s_N is a Cauchy sequence. Indeed, given $\varepsilon > 0$ choose N_ε so that $\sum_{N_\varepsilon}^{\infty} \|x_i\| < \varepsilon$. Then one has for all $n > m \ge N_\varepsilon$

$$\|s_m - s_n\| = \left\| \sum_{i=m+1}^{n} x_i \right\| \le \sum_{m+1}^{n} \|x_i\| < \varepsilon.$$

Conversely, suppose every absolutely convergent series in X converges. To establish completeness of X we let $\{x_i : i \ge 1\}$ be a Cauchy sequence in X, and we look for a limit of this sequence. It suffices to find a convergent subsequence x_{n_k}, for if a Cauchy sequence has a convergent subsequence then the whole sequence must converge.

Since x_i is a Cauchy sequence there exist $n_1 < n_2 < \dots$ such that

$$\|x_m - x_{n_k}\| \le 2^{-k} \text{ for all } m \ge n_k \text{ and } k \ge 1.$$

The series $\sum y_i$ with $y_1 = x_{n_1}$ and $y_k = x_{n_{k+1}} - x_{n_k}$ is then absolutely convergent since

$$\|y_k\| = \|x_{n_{k+1}} - x_{n_k}\| \le 2^{-k}.$$

The partial sums of this series are precisely the x_{n_k}, and by hypothesis they converge to some $x_* \in X$. □

Exercise 31. Use this completeness criterion to prove Theorem 6.

17. Sums of Banach spaces

If X and Y are Banach spaces, then the product space $X \times Y$ is again a vector space, with addition defined by

$$(x_1, y_1) + (x_2, y_2) \stackrel{\text{def}}{=} (x_1 + x_2, y_1 + y_2),$$

and scalar multiplication similarly. The product space is usually written as $X \oplus Y$ and called the *direct sum* of X and Y (to distinguish the set $X \times Y$ from the vector space $X \times Y$ – yes, this is pedantic.)

The direct sum of two Banach spaces can be given a norm by

$$\|(x,y)\|_{X\oplus Y} \stackrel{\text{def}}{=} \|x\|_X + \|y\|_Y.$$

With this norm $X \oplus Y$ is a Banach space.

Exercise 32. Prove that $(X \oplus Y, \|\dots\|_{X\oplus Y})$ is complete.

One can also define several other equivalent norms on the direct sum, such as $\|(x,y)\|' \stackrel{\text{def}}{=} \max\{\|x\|_X, \|y\|_Y\}$.

18. Quotients of Banach spaces

We recall that for a vector space X and a linear subspace $L \subset X$ the quotient X/L is defined to be the set of equivalence classes of the equivalence relation $x \stackrel{L}{\sim} y \Leftrightarrow x - y \in L$. If we denote the equivalence class of $x \in X$ by either $x + L$ or by $[x]_L$ then X/L is a vector space with addition and multilication defined by $[x]_L + [y]_L = [x+y]_L$, $\lambda[x]_L = [\lambda x]_L$.

If X is a normed vector space then the quantity

$$\|[x]_L\|_L \stackrel{\text{def}}{=} \inf\{\|y\| : y \stackrel{L}{\sim} x\}$$

is a seminorm on X/L.

Exercise 33. Show that $\|[x]_L\|_L$ is the distance from x to L, and that $\|[x]_L\|_L$ is indeed a seminorm.

In general $\|[x]_L\|_L$ will not be a norm: in fact $\|[x]_L\|_L$ is a norm if and only if L is a closed subspace of X.

Theorem 9. *If X is a Banach space and if L is a closed subspace then X/L with the norm $\|[x]_L\|_L$ is a Banach space.*

Function Spaces

The L^p spaces, ℓ^p, weighted L^p spaces, L^∞ and ℓ^∞, $BC(\Omega)$, Hölder continuous functions, Sobolev Spaces

Text books to look at: The books [4, 6], as well as most other books in the library with the title "Functional Analysis," describe the "classical Banach spaces" L^p, ℓ^p and $C(\Omega)$. The Hölder spaces and Sobolev spaces are easier to found in textbooks on PDE [2] or harmonic analysis.

19. The L^p spaces

In this section let $p \in [1, \infty)$ be given.

Let Ω be a set with a σ-algebra Σ and a countably additive measure $\mu : \Sigma \to [0, \infty]$. Then in 721 (1st semester real analysis) one defines $\mathcal{L}^p(\Omega, \Sigma, \mu)$ to be the set of Σ measurable functions $f : \Omega \to \mathbb{R}$ for which

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

The quantity $\| \dots \|_p$ thus defined is a seminorm, but not a norm as it vanishes on all functions $f \in \mathcal{N}$, where $\mathcal{N}$ is the set of functions which vanish almost everywhere. It is shown in 721 that $\mathcal{N}$ is a linear subspace of $\mathcal{L}^p$, and one defines $L^p = \mathcal{L}^p / \mathcal{N}$.

In the standard "abuse of language" we agree to forget to distinguish between a measurable function f and the equivalence class of functions g which coincide a.e. with f.

The quantity $\|f\|_p$ does not depend on which measurable function f one chooses to represent f, and thus defines a norm on L^p. The proof of the triangle inequality is not totally trivial but was given in 721. It was also shown in 721 that L^p is complete, so that L^p is a Banach space.

19.1. Special case – the sequence spaces ℓ^p.

If one chooses Ω to be the finite set $\Omega = \{1, \dots, N\}$ and defines μ to be the "counting measure", i.e. $\mu(E)$ is the number of elements of $E \subset \{1, \dots, N\}$, then L^p is a finite (N) dimensional space with norm

$$\|(x_1, \dots, x_N)\|_p = \{|x_1|^p + \dots + |x_N|^p\}^{1/p}.$$

If one sets $\Omega = \mathbb{N}$, and lets μ again be the counting measure, then L^p is the space of sequences $\{x_i : i \in \mathbb{N}\}$ for which

$$\|(x_i)_{i\geq 1}\|_p = \{\textstyle\sum_{i=1}^{\infty} |x_i|^p\}^{1/p} < \infty.$$

This space is denoted by ℓ^p, or $\ell^p(\mathbb{N})$.

One can also consider the space of bi-infinite sequences $(x_i)_{i\in\mathbb{Z}}$ whose p-norm $\sum_{i\in\mathbb{Z}} |x_i|^p$ is finite. This space is denoted by $\ell^p(\mathbb{Z})$.

19.2. Special case – weighted L^p spaces.

Let Ω be an open subset of $\mathbb{R}^n$, and let $w : \Omega \to (0, \infty)$ be a positive measurable function. We let μ be the measure $d\mu = w(x)dx$, i.e. for any measurable $E \subset \Omega$ we put

$$\mu(E) \stackrel{\text{def}}{=} \int_E w(x)\ dx.$$

Then $L^p(\Omega, w(x)dx)$ is the space of measurable functions for which

$$\|f\|_{L^p(\Omega,w(x)dx)} = \left\{ \int_\Omega w(x)|f(x)|^p\ dx \right\}^{1/p}$$

is finite.

The choice $w(x) = 1$ gives us the space commonly denoted by $L^p(\Omega)$.

20. L^∞ and ℓ^∞

We have introduced the L^p spaces for $1 \leq p < \infty$. It is natural to extend the definition to $p = \infty$ by defining the essential supremum of a measurable function $f : \Omega \to \mathbb{R}$ by

$$\operatorname*{ess.sup}_{x\in\Omega} f(x) \stackrel{\text{def}}{=} \inf\{M \in \mathbb{R} : \mu\{x : f(x) \geq M\} = 0\}.$$

The L^∞ norm is then defined to be

$$\|f\|_\infty \stackrel{\text{def}}{=} \operatorname*{ess.sup}_{x\in\Omega} |f(x)|.$$

If we again agree to identify functions which coincide except on a set of μ-measure zero then

$$L^\infty(\Omega, \Sigma, \mu) \stackrel{\text{def}}{=} \{f : \|f\|_\infty < \infty\}$$

is a Banach space. Elements of L^∞ are called *essentially bounded* functions.

If one chooses $\Omega = \mathbb{N}$ or $\Omega = \mathbb{Z}$ then the corresponding L^∞ space is a space of bounded sequences $(x_i)_{i\in\mathbb{N}}$ (or $(x_i)_{i\in\mathbb{Z}}$ respectively) with the supremum norm

$$\|(x_i)\|_\infty = \sup_i |x_i|.$$

The resulting sequence space is denoted by $\ell^\infty(\mathbb{N})$ or $\ell^\infty(\mathbb{Z})$.

21. Hölder's inequality.

For $f \in L^p$ and $g \in L^q$ where p and q are conjugate exponents, meaning

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

the product fg is integrable and one has

$$\|fg\|_1 \leq \|f\|_p \, \|g\|_q$$

i.e.

$$\int_\Omega |f(x)g(x)| \, d\mu \leq \left\{ \int_\Omega |f(x)|^p \, d\mu \right\}^{1/p} \left\{ \int_\Omega |g(x)|^q \, d\mu \right\}^{1/q}.$$

(When $p = 1$ one has $q = \infty$. The inequality still holds provided $\|g\|_\infty$ is properly interpreted – see below.)

The following statement shows that Hölder's inequality is in a sense optimal. It also provides a useful description of the L^p norm.

Lemma 10. *Let $1 \leq p \leq \infty$.*

For any $f \in L^p(\Omega, \Sigma, \mu)$ one has

$$(\dagger) \qquad \|f\|_p = \sup_{\|g\|_q \leq 1} \int_\Omega f(x)g(x)d\mu(x).$$

If $p < \infty$ then the supremum is attained by the choice

$$g(x) = A|f(x)|^{p-1}\mathrm{sign}f(x),$$

where $A = \|f\|_p^{1-p}$, and where $\mathrm{sign}f(x)$ is $+1$, 0 or -1 depending on whether $f(x) > 0$, $= 0$ or < 0 respectively.

This Lemma immediately implies that the L^p norm is a seminorm. Indeed, for each fixed $g \in L^q$ the integral in the righthand side in $(\dagger)$ is linear in f and hence defines a seminorm. The supremum of these seminorms must again be a seminorm.

Proof. Hölder's inequality directly implies that $\|f\|_p$ does not exceed the supremum, and if $p < \infty$ one can substitute the given $g(x)$ to verify that $\|f\|_p$ actually equals the supremum.

If $p = \infty$ then one takes $E_\varepsilon = \{x : |f(x)| \geq \|f\|_\infty - \varepsilon\}$. By assumption $\mu(E_\varepsilon) > 0$ for all $\varepsilon > 0$ (this is essentially the definition of the ess.sup) so we can define $g_\varepsilon(x) = |E_\varepsilon|^{-1}\chi_{E_\varepsilon}(x)\mathrm{sign}f(x)$.

One then has

$$\begin{aligned} \int f(x)g_\varepsilon(x)d\mu(x) &= \frac{1}{|E_\varepsilon|} \int_{E_\varepsilon} |f(x)|d\mu(x) \\ &\geq \|f\|_\infty - \varepsilon \end{aligned}$$

with $\varepsilon > 0$ arbitrary. □

Exercise 34. Let X be the unit ball in $L^1(-1, 1)$. Does the function $F : X \to \mathbb{R}$ given by

$$F(f) \stackrel{\mathrm{def}}{=} \int_{-1}^{1} (1 - x^2) f(x) \, dx$$

attain a maximum on X?

Exercise 35. Prove that for any measurable $f : \Omega \to \mathbb{R}$ with $\|f\|_p < \infty$ for all $p < \infty$ one has

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

Exercise 36. (An interpolation inequality.) Suppose that $f \in L^{p_0}$ and $f \in L^{p_1}$ for $1 \le p_0 < p_1 < \infty$. Show that $f \in L^p$ for all $p \in [p_0, p_1]$, and that

$$\|f\|_p \le \|f\|_{p_0}^{1-\theta} \, \|f\|_{p_1}^{\theta},$$

provided $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$.

Exercise 37.

(i) Show that if $\mu(\Omega) < \infty$ then $L^p \subset L^{p'}$ if and only if $p \ge p'$.

(ii) Assume again that $\mu(\Omega) < \infty$ and let $1 \le p < \infty$. Show that $L^\infty(\Omega)$ is a dense subspace of $L^p(\Omega)$.

(iii) Show that $\ell^p(\mathbb{N}) \subset \ell^{p'}(\mathbb{N})$ if and only if $p \le p'$.

Exercise 38.

(i) Let $\Omega = \mathbb{R}^n$ and let μ be Lebesgue measure. Give an example of a function f which belongs to L^p if and only if $p = 1999$.

(ii) Give an example which belongs to all L^p with $p < \infty$, but not to L^∞.

(iii) Give an example of a function $f \in \cap_{p<\infty} L^p$ which is not essentially bounded on any open $E \subset \mathbb{R}^n$.

22. $BC(\Omega)$

Let Ω be a topological space, and let $BC(\Omega)$ be the set of bounded and continuous real valued functions on Ω. This is a Banach space with norm

$$\|f\| = \sup_{x\in\Omega} |f(x)|.$$

If Ω is compact then all continuous fuctions on Ω are bounded and we simply write $C(\Omega)$.

Exercise 39. Observe that if $\Omega = \mathbb{N}$ then $BC(\Omega) = \ell^\infty(\mathbb{N})$.

Exercise 40. Let $\Omega = \mathbb{R}^n$. Then every bounded continuous function is also an essentially bounded measurable function on $\mathbb{R}^n$.

Show that $BC(\mathbb{R}^n) \subset L^\infty(\mathbb{R}^n)$ is a closed subspace and that $BC(\mathbb{R}^n) \ne L^\infty(\mathbb{R}^n)$.

Exercise 41. If Ω is a metric space then we can distinguish uniformly continuous functions among the merely continuous functions on Ω. (Recall f is uniformly continuous if for all $\varepsilon > 0$ a $\delta_\varepsilon > 0$ exists such that $|f(x) - f(y)| \leq \varepsilon$ for all $x, y \in \Omega$ with $d(x, y) < \delta_\varepsilon$.)

(i) Show that $BUC(\Omega)$, the space of bounded uniformly continuous functions on Ω, is a closed subspace of $BC(\Omega)$.

(ii) Find a function $f \in BC(\mathbb{R})$ which does not lie in $BUC(\mathbb{R})$.

(iii) Let

$$t(x) = \begin{cases} 1 - 2|x| & \text{for } |x| \leq 1/2 \\ 0 & \text{otherwise} \end{cases}$$

be the "tent function," and consider the map F from $\ell^\infty(\mathbb{Z})$ to $BUC(\mathbb{R})$ which assigns the function

$$F((s_i)_{i\in\mathbb{Z}})(x) = \sum_{i\in\mathbb{Z}} s_i t(x - i)$$

to the sequence $(s_i)_{i\in\mathbb{Z}} \in \ell^\infty(\mathbb{Z})$.

Show that F is an isometry.

23. Hölder continuous functions

Let $\Omega \subset \mathbb{R}^n$ be open, and let $\alpha \in (0, 1]$ be a fixed constant. A continuous function $f : \Omega \to \mathbb{R}$ is said to *Hölder continuous with exponent* α if there is a constant $C < \infty$ such that for all $x, y \in \Omega$

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

In the special case $\alpha = 1$ one speaks of *Lipschitz continuous functions* rather than Hölder continuous functions.

The best constant C is given by

$$[f]_{\alpha;\Omega} = \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}$$

The space of α-Hölder continuous functions is denoted by $C^\alpha(\Omega)$.

Exercise 42. Show that $[f]_{\alpha;\Omega}$ is a seminorm on $C^\alpha(\Omega)$.

(Suggestion: For each $x \neq y$ the quantity $p_{x,y}(f) = |x - y|^{-\alpha}|f(x) - f(y)|$ is a seminorm.)

The space of α-Hölder continuous functions with norm

$$\begin{aligned} \|f\|_{C^\alpha} &= \sup_{x\in\Omega} |f(x)| + \sup_{x\neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha} \\ &= \|f\|_{L^\infty} + [f]_{\alpha;\Omega}. \end{aligned}$$

is a Banach space.

Exercise 43. Show that the norm $\|f\|_{C^\alpha}$ is complete.

Exercise 44. Let $\Omega = (-1,1) \subset \mathbb{R}$, and define $f(x) = \sqrt{|x|}$.

(i) For which $\alpha \in (0,1]$ does $f(x)$ belong to $C^\alpha(\Omega)$?

(ii) Let $g(x)$ be a continuously differentiable function on the closed interval $-1 \leq x \leq 1$. Show that

$$\|f - g\|_{C^{1/2}} \geq 1.$$

Are polynomials dense in $C^{1/2}(-1,1)$? (Compare your answer with the Stone Weierstrass theorem.)

(iii) For each $a \in (-1,1)$ define $f_a(x) = \sqrt{|x-a|}$. Show that $f_a \to f$ uniformly as $a \to 0$. Is it true that

$$\lim_{a \to 0} \|f_a - f\|_{C^{1/2}} = 0?$$

(iv) Is the space $C^{1/2}(-1,1)$ separable?

Why were Hölder continuous functions introduced?

Hölder spaces are used extensively in the study of partial differential equations. One of the first places where one encounters them is in "potential theory," where one looks for solutions $f : \mathbb{R}^3 \to \mathbb{R}$ of the Poisson equation

$$\Delta f = \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \frac{\partial^2 f}{\partial x_3^2} = \rho(x)$$

for a given function $\rho : \mathbb{R}^3 \to \mathbb{R}$. (If ρ represents the distribution of charge, then the solution to this equation represents the potential of the electric field produced by the charges.) If ρ is compactly supported and continuous then the solution is given by

$$f(x) = -\frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{\rho(y)}{|x-y|}\, dy.$$

The integral in this formula is called the *Newton potential* of the charge distribution ρ.

Instead of wondering how this formula was derived one can try to simply verify it by computing the second derivatives of the function $f(x)$ defined by the Newton potential. After differentiating under the integral one ends up with the following integrals

$$(*) \qquad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{|x-y|^2 \delta_{ij} - 3(x_i - y_i)(x_j - y_j)}{|x-y|^5}\, \rho(y)\, dy$$

where $\delta_{ij} = 0$ for $i \neq j$ and 1 if $i = j$.

Now the integrand in these integrals are bounded by $C|x-y|^{-3}$, which is not an integrable function, so that the integrals cannot be interpreted as Lebesgue integrals, and so that the differentiation under the integral which gave us $(*)$ may not even be justified.

It turns out that if ρ is merely continuous then the Newton potential need not be a twice differentiable function. However, if ρ is known to be Hölder continuous of some exponent $\alpha \in (0,1)$, then one can justify $(*)$ and the second derivatives of f turn out to exist and they even turn out to be Hölder continuous functions of the same exponent α.

24. Sobolev Spaces

Let $\Omega \subset \mathbb{R}^n$ be open. Then $W^{1,p}(\Omega)$ is the set of functions $f \in L^p(\Omega)$ whose partial derivatives in the sense of distributions are again L^p functions, i.e. the distributions $D_i f \in \mathcal{D}'(\Omega)$ are actually of the form $D_i f = g_i$ for certain $g_i \in L^p(\Omega)$.

One can formulate this without referring to distributions by saying that $f \in L^p(\Omega)$ belongs to $W^{1,p}(\Omega)$ if there are functions $g_1, \ldots, g_n \in L^p(\Omega)$ such that for all smooth compactly supported functions $\varphi : \Omega \to \mathbb{R}$ one has

$$\int_\Omega f(x) \frac{\partial \varphi}{\partial x_i}\, dx = \int_\Omega g_i(x)\varphi(x)\, dx.$$

The space $W^{1,p}(\Omega)$ can be given the following norm

$$\|f\|_{W^{1,p}} \stackrel{\text{def}}{=} \|f\|_{L^p} + \|\partial_1 f\|_{L^p} + \ldots + \|\partial_n f\|_{L^p}$$

or the equivalent norms

$$\|f\|'_{W^{1,p}} = \left\{ \int_\Omega \left(|f|^p + |\partial_1 f(x)|^p + \ldots + |\partial_n f(x)|^p\right)\, dx \right\}^{1/p}.$$

and

$$\|f\|''_{W^{1,p}} = \left\{ \int_\Omega \left(|f|^p + |\nabla f(x)|^p\right)\, dx \right\}^{1/p}.$$

Exercise 45. Verify that $W^{1,p}(\Omega)$ with the given norm is complete.

More generally one defines $W^{m,p}(\Omega)$ to be the space of functions which together with their distributional derivatives of order at most m belong to $L^p(\Omega)$. A norm on $W^{m,p}(\Omega)$ is

$$\|f\|^p_{W^{m,p}} = \sum_{0 \le |\alpha| \le m} \int_\Omega |D^\alpha f(x)|\, dx.$$

Exercise 46. Let $\Omega = \{x \in \mathbb{R}^n : |x| < 1\}$ be the open unit ball in $\mathbb{R}^n$.

(i) For which $a \in \mathbb{R}$ does the function $f(x) = |x|^a$ belong to $W^{1,p}(\Omega)$?

(ii) Assume $1 \le p < n$ and construct a function $f \in W^{1,p}(\Omega)$ which is unbounded in every open subset $\Omega' \subset \Omega$. (Suggestion: try a function of the form $\sum_{i=1}^\infty c_i |x - r_i|^{-a}$ where r_i is an enumeration of the points in Ω with rational coordinates, and a, c_i are to be chosen appropriately.)

Exercise 47. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function, and define

$$g(x) = \begin{cases} f(x) & |x| \le 1 \\ 0 & |x| > 1 \end{cases}$$

Does f belong to $W^{1,p}(\mathbb{R}^n)$ for any $p \in [1, \infty]$?

Approximation Theorems

Approximation of L^p functions by continuous functions, The convolution product, Mollification, Smoothness of the mollification, Approximation in Sobolev spaces, Approximation of Hölder continuous functions

25. Approximation of L^p functions by continuous functions

Let $\Omega \subset \mathbb{R}^n$ be an open subset, and fix some $p \in [1, \infty)$. Denote the space of compactly supported continuous functions on Ω by $C_c(\Omega)$.

Theorem 11. *$C_c(\Omega)$ is dense in $L^p(\Omega)$.*

Proof. Let $f \in L^p$ and $\varepsilon > 0$ be given. The sequence of functions

$$f_k(x) = \begin{cases} 0 & \text{if } |x| \geq k; \text{ otherwise, if } |x| < k \text{ then,} \\ k & \text{when } f(x) > k \\ f(x) & \text{when } |f(x)| \leq k \\ -k & \text{when } f(x) < -k \end{cases}$$

converges in L^p to f (use the dominated convergence theorem to verify this.)

We choose k large enough so that $\|f - f_k\|_p < \varepsilon/2$.

By Lusin's theorem from "721" (e.g. see [4, theorem 2.23]) there exists a continuous function φ whose support is contained in $\{x \in \Omega : |x| < k\}$, which satisfies $|\varphi(x)| \leq k$, and which coincides with f_k except on a set whose measure we can assume is arbitrarily small. We will assume that $E = \{x : \varphi(x) \neq f_k(x)\}$ has measure at most $(\varepsilon/4k)^p$. Then on E one has $|\varphi - f_k| \leq 2k$ and thus

$$\|\varphi - f_k\|_p \leq \left\{ \int_E (2k)^p \, dx \right\}^{1/p} = 2km(E)^{1/p} \leq \varepsilon/2.$$

The compactly supported function φ is our approximation to f. Indeed one has

$$\|f - \varphi\|_p \leq \|f - f_k\|_p + \|f_k - \varphi\|_p \leq \varepsilon.$$

□

Exercise 48. Does this proof also work if $p = \infty$? Is $C_c(\Omega)$ dense in $L^\infty(\Omega)$?

The approximation theorem we have just shown has a few drawbacks. First, the approximation is only continuous, and this can be improved. Second, the method of approximation is not "transparent" in the sense that we do not have an explicit formula for the approximation function. One consequence of this is that the approximation only applies to L^p spaces, and that it is not clear how to generalize it to Sobolev or Hölder spaces.

There is a standard method, called "mollification," of approximating rough functions which does not suffer from these defects. Briefly stated, mollification of a function f is the same as computing the convolution of f with a suitable test function φ, so we begin by collecting some facts about the convolution of functions.

26. The convolution product

The following is known as Jensen's inequality.

Lemma 12. *Let $\phi : \mathbb{R} \to \mathbb{R}$ be convex, let μ be a measure on Ω with $\mu(\Omega) = 1$, and let $g : \Omega \to \mathbb{R}$ be μ integrable. Then*

$$\phi\left(\int_\Omega g(x)\, d\mu(x)\right) \leq \int_\Omega \phi(g(x))\, d\mu(x).$$

Proof. Since ϕ is convex we have

$$\phi(s) = \sup_{t\in\mathbb{R}} \{\phi(t) + \phi'(t)(s-t)\}.$$

Thus

$$\begin{aligned}
\phi\left(\int_\Omega g(x)\, d\mu(x)\right) &= \sup_{t\in\mathbb{R}} \phi(t) + \phi'(t)\left(\int_\Omega g(x)\, d\mu(x) - t\right) \\
&= \sup_{t\in\mathbb{R}} \int_\Omega \left(\phi(t) + \phi'(t)(g(x)-t)\right)\, d\mu(x) \\
&\leq \int_\Omega \sup_{t\in\mathbb{R}} \left(\phi(t) + \phi'(t)(g(x)-t)\right)\, d\mu(x) \\
&= \int_\Omega \phi(g(x))\, d\mu(x).
\end{aligned}$$

□

For any two functions $f, g : \mathbb{R}^n \to \mathbb{R}$ one defines the convolution

$$f * g(x) = \int_{\mathbb{R}^n} f(y) g(x-y)\, dy, \tag{1}$$

provided one can make sense of the integral. This is clearly the case if, say, $f \in L^1(\mathbb{R}^n)$ and $g \in L^\infty(\mathbb{R}^n)$. More generally one has

Theorem 13. *If $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$ with $q = p/(p-1)$, then the convolution product $f * g(x)$ is well defined for each $x \in \mathbb{R}^n$, and one has*

$$|f * g(x)| \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Proof. Apply Hölder's inequality. □

Exercise 49. Show that under the same hypotheses $f * g$ is actually a continuous function on $\mathbb{R}^n$, and that

$$\lim_{|x|\to\infty} f * g(x) = 0.$$

Theorem 14. *If $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$ then $f(y)g(x-y)$ is a Lebesgue integrable function of y for almost every $x \in \mathbb{R}^n$, and the convolution product $f * g(x)$ defined in (1) is a measurable function.*

Moreover one has $f * g \in L^p$ and

$$\|f * g\|_p \leq \|f\|_1 \, \|g\|_p.$$

Proof. We assume without loss of generality that f and g are nonnegative, and that $\int f(x)\, dx = 1$.

The function $\phi(s) = |s|^p$ is convex and hence Jensen's inequality implies

$$(f * g)(x)^p \leq \int f(x) g(x-y)^p \, dy.$$

Integration over x and application of the Fubini-Tonelli theorem then gives

$$\begin{aligned}
\int_{\mathbb{R}^n} (f * g)(x)^p \, dx &\leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x) g(x-y)^p \, dy \, dx \\
&= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x) g(x-y)^p \, dx \, dy \\
&= \int_{\mathbb{R}^n} f(x) \, dx \cdot \int_{\mathbb{R}^n} g(x-y)^p \, dy \\
&= \|g\|_p^p.
\end{aligned}$$

□

Exercise 50. Prove the following "local version" of Jensen's inequality.

If $f \in L^1(\mathbb{R}^n)$ vanishes outside of the ball of radius $\varepsilon > 0$, and if $g \in L^p(\Omega)$ for some domain $\Omega \subset \mathbb{R}^n$, then the convolution $f * g(x)$ as defined in (1) exists for almost every $x \in \Omega_\varepsilon$. Moreover,

$$\|f * g\|_{L^p(\Omega_\varepsilon)} \leq \|f\|_{L^1(\mathbb{R}^n)} \|g\|_{L^p(\Omega)}.$$

Here $\Omega_\varepsilon \stackrel{\text{def}}{=} \{x \in \Omega : B(x, \varepsilon) \subset \Omega\}$.

27. Mollification

To mollify a function $f \in L^1_{\text{loc}}(\Omega)$ one needs a compactly supported smooth function $\varphi \in \mathcal{D}(\mathbb{R}^n)$ with

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

We will assume that φ is supported in the Euclidean unit ball of $\mathbb{R}^n$, and that $\varphi \geq 0$. For each $\varepsilon > 0$ we then define

$$\begin{aligned}
f_\varepsilon(x) &= \int f(y) \varepsilon^{-n} \varphi\left(\frac{x-y}{\varepsilon}\right) dy \qquad (2) \\
&= \int f(y) \varphi_\varepsilon(x-y) \, dy
\end{aligned}$$

where $\varphi_\varepsilon(x) \stackrel{\text{def}}{=} \varepsilon^{-n}\varphi(x/\varepsilon)$. Substitution $y = x - z$ or $y = x - \varepsilon w$ gives the following alternative expressions

$$\begin{aligned} f_\varepsilon(x) &= \int f(x-z)\varphi_\varepsilon(z)\, dz \\ &= \int f(x-\varepsilon w)\varphi(w)\, dw. \end{aligned}$$

The mollification $f_\varepsilon(x)$ is only defined for $x \in \Omega_\varepsilon$. Of course, for $\Omega = \mathbb{R}^n$ one has $\Omega_\varepsilon = \Omega$ so that f_ε is defined on all of $\mathbb{R}^n$.

The integral(2) is taken over Ω, but since $\varphi(x) = 0$ for $|x| \geq 1$ we may also integrate over $B(x,\varepsilon) = \{y \in \Omega : |x-y| \leq \varepsilon\}$.

The definition shows that mollification is a linear operation: for any two locally integrable f and g and any $\lambda, \mu \in \mathbb{R}$ one has

$$(\lambda f + \mu g)_\varepsilon = \lambda f_\varepsilon + \mu g_\varepsilon.$$

Theorem 15.
(i) *If $f \in C_c(\mathbb{R}^n)$ then f_ε converges uniformly to f.*
(ii) *If $f \in L^p(\mathbb{R}^n)$ then f_ε converges to f in $L^p(\mathbb{R}^n)$.*

Proof. (i) Since f is compactly supported and continuous it is uniformly continuous. So, given $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ if $|x - y| < \delta$. The mollification $f_{\delta'}$ with $0 < \delta' < \delta$ then satisfies

$$\begin{aligned} |f_{\delta'}(x) - f(x)| &\leq \left| \int \varphi_{\delta'}(x-y)(f(y) - f(x))\, dy \right| \\ &\leq \sup_{|y-x|\leq\delta'} |f(y) - f(x)| \cdot \int \varphi_{\delta'}(x-y)\, dy \\ &\leq \varepsilon. \end{aligned}$$

(ii) Let $f \in L^p(\mathbb{R}^n)$ be given. Choose a $g \in C_c(\mathbb{R}^n)$ with $\|g - f\|_p < \varepsilon/3$. The triangle inequality gives us for any $\delta > 0$

$$\|f - f_\delta\|_p \leq \|f - g\|_p + \|g - g_\delta\|_p + \|g_\delta - f_\delta\|_p. \tag{3}$$

The first term is not more than $\varepsilon/3$. The last term can be estimated by

$$\|g_\delta - f_\delta\|_p = \|\varphi_\delta * (g - f)\|_p \leq \|\varphi\|_1 \|g - f\|_p = \|g - f\|_p < \varepsilon/3.$$

If $g(x) \equiv 0$ for $|x| \geq R$ then g_δ vanishes for $|x| \geq R + \delta$, and thus the middle term in (3) is bounded by

$$\|g - g_\delta\|_p \leq \big(m\{x : |x| < R + \delta\}\big)^{1/p} \|g - g_\delta\|_\infty$$

Since $g_\delta \to g$ uniformly we can make this less than $\varepsilon/3$ by choosing δ sufficiently small.

Adding the three estimates for the respective terms in (3) we get

$$\|f - f_\delta\|_p < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

if δ is small enough. □

A similar argument using the "local version" of Jensen's inequality leads to the following.

Theorem 16.

(i) *If $f : \Omega \to \mathbb{R}$ is continuous then f_ε converges uniformly to f on any compact $K \subset \Omega$.*

(ii) *If $f \in L^p_{\mathrm{loc}}(\Omega)$ then f_ε converges in $L^p(K)$ for any compact $K \subset \Omega$.*

28. Smoothness of the mollification

Theorem 17. *Let $f \in L^1_{\mathrm{loc}}(\Omega)$ for some open $\Omega \subset \mathbb{R}^n$.*

(i) *f_ε is a smooth function on Ω_ε.*

(ii) *The operations of mollification and differentiation commute, i.e. if both f and $D_i f$ are locally integrable then $D_i(f_\varepsilon) = \big(D_i f\big)_\varepsilon$.*

The second statement in this theorem says that the interchange of differentiation and integration in

$$\frac{\partial}{\partial x_i} \int f(x-z)\varphi_\varepsilon(z)\, dz = \int \frac{\partial f(x-z)}{\partial x_i}\varphi_\varepsilon(z)\, dz$$

is allowed when the derivative is interpreted in the sense of distributions.

Proof. Using the dominated convergence theorem one can justify differentiation under the integral to obtain

$$\begin{aligned} \frac{\partial f_\varepsilon}{\partial x_i}(x) &= \lim_{t\to 0} \int f(y) \frac{\varphi_\varepsilon(x + te_i - y) - \varphi_\varepsilon(x-y)}{t}\, dy \\ &= \int f(y) \frac{\partial \varphi_\varepsilon}{\partial x_i}(x-y)\, dy. \end{aligned} \tag{$*$}$$

It follows that the mollification of any locally integrable function is in fact differentiable. By repeating this argument one also obtains all higher derivatives, and one has

$$D^\alpha f_\varepsilon(x) = \int f(y) D^\alpha \varphi_\varepsilon(x-y)\, dy.$$

To prove (ii) we observe that

$$\frac{\partial \varphi_\varepsilon(x-y)}{\partial x_i} = -\frac{\partial \varphi_\varepsilon(x-y)}{\partial y_i}$$

which allows us to rewrite $(*)$ as

$$\begin{aligned} \frac{\partial f_\varepsilon}{\partial x_i}(x) &= -\int f(y) \frac{\partial \varphi_\varepsilon}{\partial y_i}(x-y)\, dy \\ &= -\langle f, D_i \varphi_{\varepsilon,x} \rangle \\ &= \langle D_i f, \varphi_{\varepsilon,x} \rangle \end{aligned}$$

in which $\varphi_{\varepsilon,x}(y) = \varphi_\varepsilon(x-y)$ is a test function on Ω and the last line is to be interpreted in the sense of distributions. We are assuming that the distributional derivative $D_i f$ actually belongs to L^1_{loc} and so we have

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

□

29. Approximation in Sobolev spaces

If $f \in W^{m,p}(\mathbb{R}^n)$ then we have just shown that for any multi-index α with $|\alpha| \leq m$ one has

$$D^\alpha(f_\varepsilon) = (D^\alpha f)_\varepsilon,$$

and also that as $\varepsilon \searrow 0$ the smooth functions $(D^\alpha f)_\varepsilon$ converge in $L^p(\mathbb{R}^n)$ to $D^\alpha f$. This implies that f_ε converges to f in $W^{m,p}(\mathbb{R}^n)$ so that we have proved

Theorem 18. *$W^{m,p}(\mathbb{R}^n) \cap C^\infty(\mathbb{R}^n)$ is a dense subspace of $W^{m,p}(\mathbb{R}^n)$.*

For functions $f \in W^{m,p}(\Omega)$ where $\Omega \subsetneq \mathbb{R}^n$ is open, mollification does not produce a function which is defined on all of Ω, and we don't get an analogous density theorem without some extra work. Nevertheless, the following is true and I refer to [2, §5.3] for a proof.

Theorem 19. *$W^{m,p}(\Omega) \cap C^\infty(\Omega)$ is a dense subspace of $W^{m,p}(\Omega)$.*

As Evans points out [2] the smaller space $C^\infty(\bar{\Omega})$ is in general NOT dense in $W^{k,p}(\Omega)$, but it is dense if one assumes that Ω has smooth boundary.

Using mollification one does obtain smooth approximations of functions $f \in W^{k,p}(\Omega)$, but only in the following sense.

Theorem 20. *If $f \in W^{k,p}(\Omega)$, then for any $\delta > 0$ the functions $f_\varepsilon|\Omega_\delta$ converge in $W^{k,p}(\Omega_\delta)$ to $f|\Omega_\delta$.*

Exercise 51. Show that $C_c^\infty(\mathbb{R}^n)$ is dense in $W^{m,p}(\mathbb{R}^n)$.

Exercise 52. Show that if $f, g \in W^{1,2}(\mathbb{R}^n)$ then the product $h(x) = f(x)g(x)$ belongs to $W^{1,1}(\mathbb{R}^n)$, and that the product rule holds, i.e.

$$D_i h = f D_i g + g D_i f.$$

Hint: first prove this assuming $f \in C^\infty \cap W^{1,2}$, then apply an approximation theorem.

Exercise 53. Let $f \in W^{1,p}(\mathbb{R}^n)$, and assume that $\Psi : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function with $\forall_{s \in \mathbb{R}} |\Psi'(s)| \leq M$ for some finite M. Show that the composition $g(x) = \Psi(f(x))$ again belongs to $W^{1,p}(\mathbb{R}^n)$, and that the chain rule holds, i.e.

$$D_i \Psi(f(x)) = \Psi'(f(x)) D_i f(x).$$

30. Approximation of Hölder continuous functions

We have seen that smooth functions are not dense in $C^\alpha(\Omega)$ for any $\alpha \in (0,1)$ so we cannot expect f_ε to converge to f in the α Hölder norm for any $f \in C^\alpha(\mathbb{R}^n)$. However, the mollifications f_ε do converge uniformly to f and the way in which they do this turns out to give a useful description of Hölder continuity.

Theorem 21. *Let $f \in C^\alpha(\mathbb{R}^n)$. Then f_ε converges uniformly to f as $\varepsilon \searrow 0$, and one has*

$$\|f - f_\varepsilon\|_\infty \leq [f]_\alpha \varepsilon^\alpha$$
$$\|\nabla f_\varepsilon\|_\infty \leq C [f]_\alpha \varepsilon^{\alpha - 1}$$

for some constant C which only depends on the function φ in the definition of the mollification.

We recall that the seminorm $[f]_\alpha$ is defined by

$$[f]_\alpha = \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

Proof. For any $x \in \mathbb{R}^n$ one has

$$f(x) - f_\varepsilon(x) = \int \varphi_\varepsilon(x - y)(f(y) - f(x))\, dy$$

since $\int \varphi_\varepsilon = 1$. The integral is taken over the ball $|x - y| \leq \varepsilon$ on which we have by definition $|f(x) - f(y)| \leq [f]_\alpha \varepsilon^\alpha$. This gives us the estimate

$$|f(x) - f_\varepsilon(x)| \leq [f]_\alpha \varepsilon^\alpha.$$

To estimate the derivative we start with

$$\begin{aligned} D_i f_\varepsilon(x) &= \int D_i \varphi_\varepsilon(x - y) f(y)\, dy \\ &= \int D_i \varphi_\varepsilon(x - y)(f(y) - f(x))\, dy \end{aligned}$$

where we have used $\int D_i \varphi_\varepsilon(x - y)\, dy = 0$. As before the integral is over the ball $B(x, \varepsilon)$ so that we have

$$\begin{aligned} |D_i f_\varepsilon(x)| &\leq \int |D_i \varphi_\varepsilon(x - y)||f(y) - f(x)|\, dy \\ &\leq [f]_\alpha \varepsilon^\alpha \int |D_i \varphi_\varepsilon(x - y)|\, dy \\ &= C[f]_\alpha \varepsilon^{\alpha - 1} \end{aligned}$$

in which $C = \int |\nabla \varphi(z)|\, dz$, and where we have used the "scaling property"

$$\int_{\mathbb{R}^n} |\nabla \varphi_\varepsilon(x)|\, dx = \frac{1}{\varepsilon} \int_{\mathbb{R}^n} |\nabla \varphi(x)|\, dx.$$

□

This theorem tells us the following: given a Hölder continuous function f one can approximate f with smooth functions f_ε. The closer these smooth functions get to f the larger their derivatives must be. In fact one can find a function which is at most $C\varepsilon^\alpha$ away from f, and whose derivative is nowhere larger than $C/\varepsilon^{1-\alpha}$. This turns out to be a characterization of Hölder continuity.

Theorem 22. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and suppose that for any $\varepsilon > 0$ one can find a C^1 function $g_\varepsilon : \mathbb{R}^n \to \mathbb{R}$ with*

$$\|f - g_\varepsilon\|_\infty \leq A\varepsilon^\alpha, \quad \|\nabla g_\varepsilon\|_\infty \leq A\varepsilon^{\alpha - 1},$$

then f is α Hölder continuous and

$$[f]_\alpha \leq 3A.$$

Proof. For given $x, y \in \mathbb{R}^n$ we choose $\varepsilon = |x - y|$. Then

$$|f(x) - g_\varepsilon(x)| \leq A\varepsilon^\alpha, \quad |f(y) - g_\varepsilon(y)| \leq A\varepsilon^\alpha,$$

and

$$|g_\varepsilon(x) - g_\varepsilon(y)| \leq \|\nabla g_\varepsilon\|_\infty |x - y| \leq A\varepsilon^{\alpha - 1} \cdot \varepsilon, = A\varepsilon^\alpha.$$

Add these three inequalities and you get

$$\begin{aligned} |f(x) - f(y)| &\le |f(x) - g_\varepsilon(x)| + |g_\varepsilon(x) - g_\varepsilon(y)| + |f(y) - g_\varepsilon(y)| \\ &\le 3A\varepsilon^\alpha \\ &= 3A|x-y|^\alpha, \end{aligned}$$

as claimed. □

Exercise 54. Let $f \in C^\alpha([0,1])$, and for any $n \in \mathbb{N}$ let $f_n : [0,1] \to \mathbb{R}$ be the continuous function which coincides with f at $x = 0$, $\frac{1}{n}$, $\frac{2}{n}$, ... , $\frac{n-1}{n}$, and $x = 1$, and which is linear on each interval $[k/n, (k+1)/n]$.

Estimate $\|f - f_n\|_\infty$ and $\|f_n'\|_\infty$ in terms of $\varepsilon = 1/n$.

Embedding Theorems

The Sobolev embedding theorem for $p > n$, the case $1 \leq p < n$, the Isoperimetric inequality

31. The Sobolev embedding theorem for $p > n$

The Sobolev embedding theorems say that if an L^p function has a derivative in L^p then it must be more regular than just any L^p function. There are two different cases to be considered depending on the size of p, and the dividing line is at $p = n$.

Theorem 23. *If $f \in W^{1,p}(\mathbb{R}^n)$ then f is α-Hölder continuous, where $\alpha = 1 - \frac{n}{p}$, and moreover*

$$[f]_\alpha \leq C\|\nabla f\|_{L^p}$$

for some finite constant C.

The proof goes by showing that the mollifications f_ε satisfy the characterization of Hölder continuous functions from the previous chapter. We begin with the identity

Lemma 24. *If f and $D_i f$ are locally integrable then*

$$\frac{\partial f_\varepsilon(x)}{\partial \varepsilon} = -\int_{\mathbb{R}^n} D_i f(y) \varepsilon^{-n} \psi_i(\frac{x-y}{\varepsilon})\, dy, \tag{4}$$

where

$$\psi_i(x) = x_i \varphi(x).$$

Formally one can derive (4) by differentiating

$$f_\varepsilon(x) = \int f(x - \varepsilon y)\varphi(y)\, dy$$

under the integral, and substituting $y = z/\varepsilon$. It is not clear that this is allowed since f only has distributional derivatives.

Instead we argue as follows:

Proof. One has

$$\begin{aligned}\frac{\partial f_\varepsilon(x)}{\partial \varepsilon} &= \frac{\partial}{\partial \varepsilon}\int f(y)\varphi_\varepsilon(x-y)\,dy \\ &= \int f(y)\left\{-\frac{n}{\varepsilon^{n+1}}\varphi(\frac{x-y}{\varepsilon}) - \frac{x_i-y_i}{\varepsilon^{n+2}}(\partial_i\varphi)(\frac{x-y}{\varepsilon})\right\}\,dy \\ &= \int f(y)\left\{-\frac{n}{\varepsilon^{n+1}}\varphi(\frac{x-y}{\varepsilon}) - \frac{x_i-y_i}{\varepsilon^{n+1}}\frac{\partial}{\partial y_i}\varphi(\frac{x-y}{\varepsilon})\right\}\,dy \\ &= \int f(y)\frac{\partial}{\partial y_i}\left(\frac{x_i-y_i}{\varepsilon^{n+1}}\varphi(\frac{x-y}{\varepsilon})\right)\,dy\end{aligned}$$

(apply the definition of distributional derivative)

$$= -\int D_i f(y)\varepsilon^{-n}\psi_i(\frac{x-y}{\varepsilon})\,dy$$

□

Proof of the Theorem.

We show that the f_ε converge uniformly. One has

$$\left|\frac{\partial f_\varepsilon(x)}{\partial\varepsilon}\right| \le \left(\int |\nabla f(y)|^p dy\right)^{1/p} \cdot \left(\int \varepsilon^{-nq}\left|\psi_i(\frac{x-y}{\varepsilon})\right|^q dy\right)^{1/q},$$

(substitute $y = x+\varepsilon z$ in the second integral)

$$\begin{aligned}&= \|\nabla f\|_{L^p} \cdot \left(\int |\psi_i(z)|^q dz\right)^{1/q} \varepsilon^{n/q-n} \\ &= C\|\nabla f\|_{L^p}\varepsilon^{-n/p}.\end{aligned}$$

For any two $0<\varepsilon_1<\varepsilon_2<\varepsilon_*$ we therefore have

$$\begin{aligned}|f_{\varepsilon_1}(x) - f_{\varepsilon_2}(x)| &= \int_{\varepsilon_1}^{\varepsilon_2}\left|\frac{\partial f_\varepsilon(x)}{\partial\varepsilon}\right| d\varepsilon \\ &\le C\|\nabla f\|_{L^p}\int_0^{\varepsilon_*}\varepsilon^{-n/p}\,d\varepsilon \\ &\le C'\|\nabla f\|_{L^p}\varepsilon_*^{1-n/p}.\end{aligned}$$

This estimate is independent of x and we therefore see that the f_ε form a Cauchy sequence in the L^∞ norm. They must therefore converge uniformly to some continuous function. Since we already know that the f_ε converge to f in L^p, we conclude that f is continuous and that

$$\|f - f_\varepsilon\|_{L^\infty} \le C\|\nabla f\|_{L^p}\varepsilon_*^{1-n/p}. \tag{5}$$

To conclude the proof we estimate the size of ∇f_ε. We have

$$\begin{aligned}|D_i(f_\varepsilon)(x)| &= |(D_i f)_\varepsilon(x)| \\ &= \left| \int D_i f(y) \varphi_\varepsilon(x-y)\, dy \right| \\ &\leq \int |D_i f(y)||\varphi_\varepsilon(x-y)|\, dy \\ &\leq \|\nabla f\|_{L^p} \cdot \|\varphi_\varepsilon\|_{L^q} \\ &= C\|\nabla f\|_{L^p} \varepsilon^{-n/p}.\end{aligned}$$

Together with (5) this inequality implies that f is Hölder continuous of exponent $\alpha = 1 - n/p$ as claimed. □

Exercise 55. The one dimensional case of this theorem has an easier proof which is worth remembering.

Let $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function with $f'(x) \in L^p(\mathbb{R})$. Prove that

$$|f(x) - f(y)| \leq \|f'\|_{L^p} |x-y|^{1-1/p}$$

by applying Hölder's inequality to

$$f(x) - f(y) = \int_x^y f'(\xi)\, d\xi.$$

Exercise 56. Every $f \in W^{1,p}$ is $(1-n/p)$-Hölder continuous, but not every $(1-n/p)$-Hölder continuous function belongs to $W^{1,p}(\mathbb{R}^n)$. Illustrate this by giving an example of a compactly supported function $f : \mathbb{R} \to \mathbb{R}$ which is Hölder continuous of exponent $1/2$, but which does not belong to $W^{1,2}(\mathbb{R})$.

Exercise 57. Show that the unbounded function $f(x) = \log\log\left(1 + \frac{1}{|x|}\right)$ belongs to $W^{1,n}(\Omega)$.

The analogous theorem which one would expect for general domains $\Omega \subset \mathbb{R}^n$, namely that any $f \in W^{1,p}(\Omega)$ is Hölder continuous with exponent $\alpha = 1 - n/p$ is unfortunately not true. It turns out that one must impose some regularity on the boundary of the domain. Without proof I state the following;

Theorem 25. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain with smooth boundary $\partial\Omega$. Then any $f \in W^{1,p}(\Omega)$ is Hölder continuous on Ω of exponent $1 - n/p$.*

A proof can be found in [2].

32. The Sobolev inequality in the case $1 \leq p < n$

Theorem 26. *If $f \in W^{1,1}(\mathbb{R}^n)$ then $f \in L^{n/(n-1)}$ and*

$$\|f\|_{L^{n/(n-1)}} \leq \sqrt[n]{\prod_{i=1}^n \|D_i f\|_{L^1}} \leq \int_{\mathbb{R}^n} |\nabla f(x)|\, dx. \tag{6}$$

When $n = 1$ one must interpret $n/(n-1)$ as ∞.

The following proof can be found in many places. I follow Stein's exposition [7, page 129] closely.

Proof. We will assume that f is actually smooth and compactly supported, and prove the stated inequality. The general case is then obtained by using an approximation theorem.

One proves the inequality (6) by induction on the dimension n.

The case $n = 1$ is clear in view of the identity

$$f(x) = \int_{-\infty}^{x} f'(\xi)\, d\xi$$

which implies $\sup |f(x)| \leq \|f'\|_{L^1}$.

Let $n > 1$ be given and assume the inequality has been proven for all dimensions less than n. If $x \in \mathbb{R}^n$ then we write x_1 for the first coordinate of x, and $x' = (x_2, \ldots, x_n)$. We introduce the following functions

$$I_1(x_2, \ldots, x_n) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \left| \frac{\partial f}{\partial x_1}(x_1, \ldots, x_n) \right| \, dx_1$$
$$I_j(x_1) \stackrel{\text{def}}{=} \int_{\mathbb{R}^{n-1}} \left| \frac{\partial f}{\partial x_j}(x_1, \ldots, x_n) \right| \, dx'$$

For each fixed x_1 we may regard $f(x_1, x')$ as a compactly supported smooth function of $x' \in \mathbb{R}^{n-1}$ and thus the induction hypothesis tells us that

$$\left(\int_{\mathbb{R}^{n-1}} |f(x_1, x')|^{\frac{n-1}{n-2}} dx' \right)^{\frac{n-2}{n-1}} \leq (I_2(x_1) I_3(x_1) \ldots I_n(x_1))^{\frac{1}{n-1}} . \tag{7}$$

On the other hand we also have

$$|f(x_1, x')| \leq I_1(x')$$

(this is the one dimensional case applied to the function $x_1 \mapsto f(x_1, x')$). Since $\frac{n}{n-1} = 1 + \frac{1}{n-1}$ this implies

$$|f(x)|^{\frac{n}{n-1}} \leq (I_1(x'))^{\frac{1}{n-1}} |f(x)|.$$

Integration and Hölder's inequality now give us

$$\begin{aligned} &\int_{\mathbb{R}^{n-1}} |f(x_1, x')|^{\frac{n}{n-1}} \, dx' \\ &\qquad \leq \int_{\mathbb{R}^{n-1}} (I_1(x'))^{\frac{1}{n-1}} |f(x_1, x')| \, dx' \\ &\qquad \leq \left(\int_{\mathbb{R}^{n-1}} I_1(x') dx' \right)^{\frac{1}{n-1}} \left(\int_{\mathbb{R}^{n-1}} |f(x_1, x')|^{\frac{n-1}{n-2}} dx' \right)^{\frac{n-2}{n-1}} \end{aligned}$$

(substitute (7))

$$\leq \left(\int_{\mathbb{R}^{n-1}} I_1(x') dx' \right)^{\frac{1}{n-1}} (I_2(x_1) I_3(x_1) \ldots I_n(x_1))^{\frac{1}{n-1}} .$$

This holds for each $x_1 \in \mathbb{R}$. The first factor here is a constant, namely

$$\left(\int_{\mathbb{R}^{n-1}} I_1(x') dx' \right)^{\frac{1}{n-1}} = \|D_1 f\|_{L^1}^{1/(n-1)},$$

the others depend on x_1. If we integrate these use Hölder's inequality again, we get

$$\begin{aligned}\int_{-\infty}^{\infty} (I_2(x_1)I_3(x_1)\dots I_n(x_1))^{\frac{1}{n-1}}\,dx_1 &\le \prod_{j=2}^{n}\left(\int_{-\infty}^{\infty} I_j(x_1)dx_1\right)^{\frac{1}{n-1}}\\ &= \prod_{j=2}^{n}\|D_jf\|_{L^1}^{1/(n-1)}.\end{aligned}$$

Finally we get

$$\int_{\mathbb{R}^n} |f|^{\frac{n}{n-1}}\,dx \le \left(\prod_{i=1}^{n}\|D_if\|_{L^1}\right)^{\frac{1}{n-1}},$$

as promised. □

The general case $1 < p < n$ now follows.

Theorem 27. *If $f \in W^{1,p}(\mathbb{R}^n)$ with $1 < p < n$ then $f \in L^r(\mathbb{R}^n)$ where $\frac{1}{r} = \frac{1}{p} - \frac{1}{n}$. One has*

$$\|f\|_{L^t} \le \frac{n-1}{n-p}p\|\nabla f\|_{L^p}.$$

Proof. We assume that f is smooth and has compact support, and apply the $p = 1$ version to

$$g = |f|^{\frac{n-1}{n-p}p}.$$

One has

$$\begin{aligned}\int |f|^{\frac{np}{n-p}}\,dx &= \int |g|^{\frac{n}{n-1}}\,dx\\ &\le \left(\int |\nabla g|\,dx\right)^{\frac{n}{n-1}}\\ &\le \left(\frac{n-1}{n-p}p\int |f|^{\frac{n-1}{n-p}p-1}|\nabla f|\,dx\right)^{\frac{n}{n-1}}\end{aligned}$$

Since

$$\frac{n-1}{n-p}p - 1 = \frac{np-p-n+p}{n-p} = \frac{n(p-1)}{n-p}$$

we then get, by Hölder's inequlity,

$$\begin{aligned}\left(\int |f|^{\frac{np}{n-p}}\,dx\right)^{\frac{n-1}{n}} &\le \frac{n-1}{n-p}p\int |f|^{\frac{n(p-1)}{n-p}}|\nabla f|\,dx\\ &\le \frac{n-1}{n-p}p\left(\int |f|^{\frac{np}{n-p}}\,dx\right)^{\frac{p-1}{p}}\left(\int |\nabla f|^p\,dx\right)^{\frac{1}{p}}\end{aligned}$$

which after some manipulation yields the stated inequality. □

33. The isoperimetric inequality

Let $\Omega \subset \mathbb{R}^n$ by a bounded domain whose boundary is smooth. We call the $(n-1)$ dimensional measure of its boundary the *perimiter of* Ω,written $\operatorname{Per}\Omega$ (for $\Omega \subset \mathbb{R}^2$ $\operatorname{Per}\Omega$ is the length of the boundary of Ω; for $\Omega \subset \mathbb{R}^3$ $\operatorname{Per}\Omega$ is the surface area of $\partial\Omega$, etc ...)

The dimensionless quantity

$$\frac{(\operatorname{Per}\Omega)^{\frac{n}{n-1}}}{\operatorname{Vol}\Omega}$$

is called the *isoperimetric ratio* of Ω. The isoperimetric inequality states that the isoperimetric ratio of any bounded domain with smooth boundary is strictly greater than the isopermetric ratio of the unit ball in $\mathbb{R}^n$, unles Ω itself is a ball $B(x,R)$. For instance, for plane domains Ω the length L of the boundary and area A of the domain satisfy

$$\frac{L^2}{A} \geq 4\pi,$$

and that the only domains which actually attain this minimum value are circular discs. See [1, chapter 2, §10] for a proof.

It turns out that a weaker version of the isoperimetric inequality follows from the Sobolev inequality.

Theorem 28. *For any bounded domain Ω with smooth boundary $\partial\Omega$ one has*

$$|\Omega| \leq (\operatorname{Per}\Omega)^{\frac{n}{n-1}}. \tag{8}$$

Proof. For small $h > 0$ we consider the function $f_h : \mathbb{R}^n \to \mathbb{R}$ given by

$$f_h(x) = \begin{cases} 1 & (x \in \Omega) \\ \frac{1}{h}\operatorname{dist}(x,\Omega) & (\operatorname{dist}(x,\Omega) < h) \\ 0 & (\operatorname{dist}(x,\Omega) \geq h) \end{cases}$$

Then f_h converges monotonically to χ_Ω so

$$\lim_{h \searrow 0} \int |f_h(x)|^{\frac{n}{n-1}}\, dx = |\Omega|.$$

On the other hand

$$|\nabla f_h(x)| = \begin{cases} \frac{1}{h} & \text{for } 0 < \operatorname{dist}(x,\Omega) < h \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\int |\nabla f_h(x)|\, dx = \frac{1}{h}|\{x : 0 < \operatorname{dist}(x,\Omega) < h\}|.$$

which converges to the $n-1$ dimensional measure of $\partial\Omega$[1]. Sobolev's inequality applied to f_h then implies (8). □

Exercise 58. For $a \in \mathbb{R}^n$ with $a_i > 0$ let $\mathcal{R}(a)$ be the rectangle $(0,a_1) \times \ldots \times (0,a_n)$. Which rectangle with $\operatorname{Vol}(\mathrm{R}(a)) = 1$ has smallest perimiter, and compute this minimal perimiter.

[1] One may take this as definition of the perimiter of Ω

Compactness theorems

About compactness, Compact subsets of $C(K)$, The Rellich-Kondrachov theorem

Text books to look at: Rudin [5, appendix] has a summary of the facts from point-set topology about compactness, as well as a proof of the Ascoli-Arzela theorem. The Rellich-Kondrachov theorem is proven in Evans [2].

34. About Compactness

We first recall some notions from point-set topology.

A Hausdorff topological space X is *compact* if every open cover of X has a finite subcover. The space X is *sequentially compact* if every sequence in X has a convergent subsequence.

For metric spaces (X, d) compactness and sequential compactness are equivalent.

There is a third characterization of compactness, namely, a complete metric space (X, d) is compact if and only if it is *totally bounded*. By definition, a metric space (X, d) is totally bounded if for every $\varepsilon > 0$ one can find a finite number of points $x_1, \ldots, x_N \in X$ such that $X = \cup_{i=1}^{N} B(x_i, \varepsilon)$.

A subset A of a topological space X is called *precompact* if the closure of A in X is compact. If X is a metric space then $A \subset X$ is precompact in X if every sequence in A has a convergent subsequence (whose limit may or may not lie in A).

For instance, all bounded subsets of $\mathbb{R}^n$ are precompact, but only the bounded and closed subsets are compact.

A subset A of a metric space is precompact in X if and only if it is totally bounded.

35. Compact subsets of $C(K)$

If K is a compact metric space, then the space $C(K)$ of continuous functions on K is a complete metric space and the Ascoli-Arzela theorem characterizes which subsets of $C(K)$ are compact.

Let $A \subset C(K)$ be any family of functions. Then A is said to be *equicontinuous* if for any $\varepsilon > 0$ there is a $\delta > 0$ such that

$$\forall_{x,y \in K} \, \forall_{f \in A} \;\; d(x, y) < \delta \Rightarrow |f(x) - f(y)| < \varepsilon.$$

Theorem 29 (Ascoli-Arzela)**.** *A subset $A \subset C(K)$ is compact if and only if it is bounded, closed and equicontinuous.*

The proof will be given in class.

Exercise 59. Let A be the unit ball of $C^\alpha(K)$. Then A is a subset of $C(K)$. Show that A is compact.

Exercise 60.
(i) Is $A_1 = \{\sin nx : n \in \mathbb{N}\}$ a precompact subset of $C([0,1])$? Is A_1 compact?
(ii) Is A_1 as above a precompact subset of $L^1(0,1)$? Compact?
(iii) Is $A_2 = \{\frac{1}{n}\sin nx : n \in \mathbb{N}\}$ a precompact or compact subset of $C([0,1])$?
(iv) Let A_3 be the set of continuously differentiable functions $f : [0,1] \to \mathbb{R}$ with $\sup_{0\le x\le 1} |f'(x)| \le 1$. Is A_3 a (pre)compact subset of $C([0,1])$?

Exercise 61.
(i) Let $f_n \in L^\infty(\Omega) \cap L^1(\Omega)$ be a sequence of functions with $|f_n(x)| \le M$ for all $n \in \mathbb{N}$, $x \in \Omega$. Suppose f_n converges in L^1 and show that f_n converges in L^p for any $1 \le p < \infty$.
(ii) Let $\Omega = \mathbb{R}$. Find an example of a sequence of functions f_n as described above which does not converge in $L^\infty(\mathbb{R})$.
(iii) Let $f_n \in L^\infty(\Omega) \cap L^1(\Omega)$ be a sequence of functions with $\|f_n\|_{L^1} \le M$ for all $n \in \mathbb{N}$. Suppose f_n converges in L^∞ and show that f_n converges in L^p for any $1 < p \le \infty$.
(iv) Let $\Omega = \mathbb{R}$. Find an example of a sequence of functions f_n as described above in (iii) which does not converge in $L^1(\mathbb{R})$.
(v) Let $A \subset L^\infty(\Omega) \cap L^1(\Omega)$ be a family of functions which is bounded in L^∞ and precompact in L^1. Show that $A \subset L^p(\Omega)$ and that A is precompact in $L^p(\Omega)$.

36. The Rellich-Kondrachov theorem

Theorem 30. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Then any sequence of functions $f_i \in W^{1,p}(\mathbb{R}^n)$ with*

$$\operatorname{supp} f_i \subset \Omega$$

and

$$\sup_{i\in\mathbb{N}} \|f_i\|_{W^{1,p}} < \infty$$

has a subsequence which converges in $L^p(\Omega)$.

Proof. For $0 < \varepsilon < 1$ we define

$$f_{i,\varepsilon} = f * \varphi_\varepsilon.$$

If $\Omega \subset B(0,R)$ then the $f_{i,\varepsilon}$ are supported in $B(0,R+1)$ for all i and $\varepsilon \in (0,1]$.

The $f_{i,\varepsilon}$ are differentiable, and one has $\nabla f_{i,\varepsilon} = (\nabla f_i) * \varphi_\varepsilon$ so that

$$\|\nabla f_{i,\varepsilon}\|_\infty \le \|\nabla f_i\|_{L^p} \|\varphi_\varepsilon\|_{L^q} < C(\varepsilon)$$

for some $C(\varepsilon)$ which does not depend on ε.

By the mean value theorem one has

$$|f_{i,\varepsilon}(x) - f_{i,\varepsilon}(y)| \le C(\varepsilon)|x-y|, \tag{9}$$

so that the mollifications $f_{i,\varepsilon}$ are equicontinuous.

Since each $f_{i,\varepsilon}$ vanishes outside of $B(0,R+1)$ (9) implies that

$$|f_{i,\varepsilon}(x) \leq 2C(\varepsilon)(R+1). \tag{10}$$

Together (9),(10) imply that the $f_{i,\varepsilon}$ are uniformly bounded and equicontinuous, so the Ascoli-Arzela theorem guarantees the existence of a convergent subsequence $\{f_{i_j,\varepsilon} : j \in \mathbb{N}\}$ *for any fixed* $\varepsilon > 0$. We will now apply Cantor's diagonal trick to produce a subsequence which converges for $\varepsilon = (1/2)^k$ for all $k \in \mathbb{N}$ at the same time.

Cantor's argument goes like this: First choose a subsequence

$$i_1^{(1)} < i_2^{(1)} < i_3^{(1)} < \cdots$$

of the integers such that $f_{i_j^{(1)},1/2}$ converges uniformly. Next, extract a subsequence $\{i_j^{(2)} : j \in \mathbb{N}\}$ from our first sequence $\{i_j^{(1)} : j \in \mathbb{N}\}$ such that $f_{i_{j,(1/2)^j}^{(2)}}$ also converges uniformly.

Continuing inductively one finds a sequence of sequences $i_j^{(k)}$, where each $\{i_j^{(k)} : j \in \mathbb{N}\}$ is a subsequence of $\{i_j^{(k-1)} : j \in \mathbb{N}\}$, and where for each $l = 1, 2, \ldots, k$, $f_{i_j^{(k)},(1/2)^l}$ converges uniformly as $j \to \infty$.

We now define $m_k = i_k^{(k)}$. Then $f_{m_k,(1/2)^j}$ converges uniformly for every $j \in \mathbb{N}$, since for each k the "tail" $\{m_l : l \geq k\}$ is a subsequence of $n_j^{(k)}$.

Since the $f_{i,\varepsilon}$ have their support in a bounded set, it follows that for each j the $f_{m_k,(1/2)^j}$ also converge in $L^p(\mathbb{R}^n)$.

We now estimate the L^p norm of $f_i - f_{i,\varepsilon}$. By Lemma 24 we have

$$\frac{\partial f_{i,\varepsilon}}{\partial \varepsilon} = \sum_{\ell=1}^{n} \psi_{\ell,\varepsilon} * D_\ell f_i \text{ where } \psi_{i,\varepsilon}(x) = \frac{x_i}{\varepsilon^n}\varphi(\frac{x}{\varepsilon})$$

(see (4).) Hence it follows that

$$\left\| \frac{\partial f_{i,\varepsilon}}{\partial \varepsilon} \right\|_{L^p} \leq n\|\psi_{i,\varepsilon}\|_{L^1}\|D_i f_i\|_{L^p} \leq C$$

where C does not depend on n or ε. Integration from 0 to ε gives

$$\|f_i - f_{i,\varepsilon}\|_{L^p} \leq \int_0^\varepsilon \left\| \frac{\partial f_{i,\varepsilon}}{\partial \varepsilon} \right\|_{L^p} d\varepsilon' \leq C\varepsilon. \tag{11}$$

We conclude by proving that the subsequence f_{m_k} we had found above is a Cauchy sequence in L^p. To see why this is true let $\sigma > 0$ be given, and set $\varepsilon_* = \sigma/3C$ so $\|f_i - f_{i,\varepsilon}\|_{L^p} < \frac{\sigma}{3}$ for all i and all $\varepsilon \in (0, \varepsilon_*)$. Choose some $j \in \mathbb{N}$ with $(1/2)^j < \varepsilon_*$. Since the mollified sequence $f_{m_k,(1/2)^j}$ converges in L^p there is an $N < \infty$ such that

$$\|f_{m_k,(1/2)^j} - f_{m_l,(1/2)^j}\|_{L^p} < \frac{\sigma}{3}$$

for all $k, l > N$. By the triangle inequality we then get

$$\begin{aligned} \|f_{m_k} - f_{m_l}\|_{L^p} &\leq \|f_{m_k} - f_{m_k,(1/2)^j}\|_{L^p} + \\ &\quad + \|f_{m_k,(1/2)^j} - f_{m_l,(1/2)^j}\|_{L^p} + \|f_{m_l,(1/2)^j} - f_{m_l}\|_{L^p} \\ &< \frac{\sigma}{3} + \frac{\sigma}{3} + \frac{\sigma}{3} \\ &= \sigma \end{aligned}$$

for all $k, l > N$. □

Exercise 62. Justify (11), i.e. prove that if $f(x, \varepsilon)$ is a continuously differentiable function in ε, then

$$\|f(\cdot, \varepsilon_1) - f(\cdot, \varepsilon_2)\|_{L^p} \leq \int_{\varepsilon_1}^{\varepsilon_2} \left\| \frac{\partial f}{\partial \varepsilon} \right\|_{L^p} d\varepsilon.$$

Observe that this can be seen as a "continuous version" of the triangle inequality. (Hint: use lemma 10.)

Boundary values

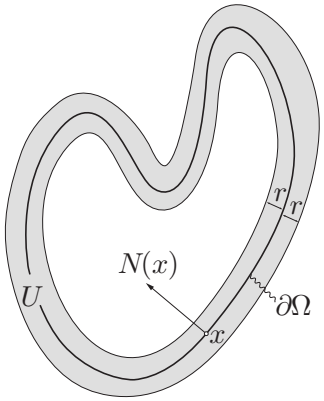
Some geometry, A trace theorem, Domains with general boundary, the space $W_o^{1,p}(\Omega)$.

Text books to look at: The material in this section is covered in more detail in Evans' [2] in the section on "trace theorems."

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. If $f \in L^p(\Omega)$ then it is meaningless to speak of the value of f at any particular point, or even of the restriction of f to any subset $E \subset \Omega$ of measure zero. If the function has distributional derivatives in $D_i f \in L^1(\Omega)$ then it turns out that one can define $f|\partial\Omega$ if $\partial\Omega$ is smooth enough.

37. Some geometry

Let $\partial\Omega$ be C^2. On $\partial\Omega$ we have a measure, namely $(n-1)$ dimensional surface measure, which we will denote by dS. The space $L^p(\partial\Omega, dS)$ is thus well defined.



At each $x \in \partial\Omega$ one can define the (inward pointing) unit normal $N(x)$. In differential geometry it is shown that the map $\Phi : \partial\Omega \times \mathbb{R} \to \mathbb{R}^n$ given by

$$\Phi(x, s) = x + sN(x)$$

is differentiable, and is one-to-one on $\partial\Omega \times (-r, r)$ for some small enough $r > 0$. The image

$$U = \Phi\left(\partial\Omega \times (-r, r)\right)$$

is an open ("tubular") neighbourhood of $\partial\Omega$.

Moreover the measures $ds\ dS$ and Lebesgue measure are "comparable" on U, by which we mean

Lemma 31. *For any nonnegative $f \in L^1(U)$ one has*

$$c\int_U f(x)\ dx \le \int_{\partial\Omega}\int_{-r}^{r} f(x+sN(x))\ ds\ dS \le C\int_U f(x)\ dx$$

for constants $0 < c < C < \infty$ which only depend on the domain Ω.

38. A trace theorem

Theorem 32. *Let $f \in C^1(\bar\Omega)$. Then there is a constant $C < \infty$ which only depends on the domain Ω, but not on f, such that*

$$\|f\,\big|_{L^p(\partial\Omega,dS)}\,\| \le C\|f\|_{W^{1,p}(\Omega)}.$$

Proof. Choose a nonincreasing smooth function $\zeta : \mathbb{R} \to \mathbb{R}$ which satisfies $\zeta(t) = 1$ for $t \le r/3$ and $\zeta(t) = 0$ for $t \ge 2r/3$. Then one has for each $x \in \partial\Omega$

$$\begin{aligned} f(x) &= -\int_0^r \frac{d\zeta(s)f(x+sN(x))}{ds}\ ds \\ &= -\int_0^r \{\zeta'(s)f(x+sN(x)) + \zeta(s)N(x)\cdot(\nabla f)(x+sN(x))\}\ ds \end{aligned}$$

Using Jensen's inequality one then gets

$$\int_{\partial\Omega}|f(x)|^p\ dS \le C\int_{\partial\Omega}\int_0^r \{|f(x+sN(x))|^p + |\nabla f(x+sN(x))|^p\}\ ds\ dS.$$

By Lemma 31 we thus get

$$\int_{\partial\Omega}|f(x)|^p\ dS \le C'\int_{U\cap\Omega}\{|f|^p + |\nabla f|^p\}\ dx \le C''\|f\|^p_{W^{1,p}(\Omega)}.$$

□

If $f \in W^{1,p}(\Omega)$ is not smooth then we can approximate f by smooth functions and thus define the restriction of f to $\partial\Omega$. In detail, given $f \in W^{1,p}(\Omega)$ we choose a sequence of functions $f_i \in C^1(\bar\Omega)$ which converges in $W^{1,p}(\Omega)$ to f. For each f_i restriction to $\partial\Omega$ gives a C^1 function on $\partial\Omega$. Theorem 32 tells us that

$$\|f_i|_{\partial\Omega} - f_j|_{\partial\Omega}\|_{L^p(\partial\Omega)} \le C\|f_i - f_j\|_{W^{1,p}(\Omega)}.$$

Therefore, since the f_i form a Cauchy sequence in $W^{1,p}(\Omega)$, their restrictions to $\partial\Omega$ form a Cauchy sequence in $L^p(\Omega)$. We call the limit of the $f_i|_{\partial\Omega}$ the restriction of f to $\partial\Omega$ or the *trace of f on $\partial\Omega$*, i.e.

$$f|_{\partial\Omega} \stackrel{\text{def}}{=} \lim_{i\to\infty} f_i|_{\partial\Omega} \in L^p(\partial\Omega).$$

Exercise 63. The trace as defined here might depend on the particular sequence $f_i \in C^1(\bar\Omega)$ one chooses to approximate f. Show that if one has two sequences $f_i, g_i \in C^1(\bar\Omega)$ which both converge in $W^{1,p}(\Omega)$ to f, then their restrictions $f_i|_{\partial\Omega}$ and $g_i|_{\partial\Omega}$ converge to the same function in $L^p(\partial\Omega)$.

Thus the restriction of any $f \in W^{1,p}(\Omega)$ is a well defined function in $L^p(\partial\Omega)$. On the other hand it cannot be true that every $f_* \in L^p(\partial\Omega)$ is the restriction of

some $f \in W^{1,p}(\Omega)$. Indeed, if $p > n$ then the Sobolev embedding theorem tells us that any $f \in W^{1,p}(\Omega)$ is continuous, so that its restriction to $\partial\Omega$ should also be continuous and therefore cannot be just any L^p function on $\partial\Omega$.

This raises the question *Which $f_* \in L^p(\partial\Omega)$ are of the form $f_* = f|_{\partial\Omega}$ for some $f \in W^{1,p}(\Omega)$?* The answer involves functions with "fractional derivatives" is beyond the scope of these notes.

39. Domains with general boundary

There is a different approach to defining the "boundary values of a function $f \in W^{1,p}(\Omega)$" which has the advantage that it makes no assumptions about the regularity of $\partial\Omega$. The trick is to abandon a direct definition of the boundary values of an $f \in W^{1,p}(\Omega)$ and merely to define when two functions $f, g \in W^{1,p}(\Omega)$ "have the same boundary values."

Let $\Omega \subset \mathbb{R}^n$ be open, not necessarily bounded. We define

$$W_o^{1,p}(\Omega) = \text{Closure of } \mathcal{D}(\Omega) \text{ in } W^{1,p}(\Omega)$$

and say that $f, g \in W^{1,p}(\Omega)$ *have the same boundary values if* $f - g \in W_o^{1,p}(\Omega)$.

That this use of the term "boundary value" is consistent with that introduced in the previous section is the main content of the following lemma.

Lemma 33. *If $\partial\Omega$ is smooth (C^2) and if $f \in W^{1,p}(\Omega)$ then $f \in W_o^{1,p}(\Omega)$ if and only if the trace $f|_{\partial\Omega}$ of f on $\partial\Omega$ vanishes.*

A proof is given in [2].

Exercise 64. Let $\Omega \subset \mathbb{R}^2$ be the open unit disc with the line segment $\{(x, 0) : 0 \le x < 1\}$ removed. Consider the function $f(x, y) = r\theta$, where $r > 0$ and $\theta \in (0, 2\pi)$ are polar coordinates of (x, y).

Show that $f \in W^{1,p}(\Omega)$ for any $p \le \infty$.

Compute

$$\lim_{\substack{P \to Q \\ P \in \Omega}} f(P)$$

for any $Q \in \partial\Omega$. Discuss what boundary values f has, and which of the definitions given above apply in this situation?

Exercise 65. Let $\Omega = \mathbb{R}_+$, and show that for any $f \in W_o^{1,2}(\Omega)$ one has $\dfrac{f(x)}{x} \in L^2(\Omega)$, as well as the following inequality

$$\left\| \frac{f(x)}{x} \right\|_{L^2(\Omega)} \le 2 \|f'(x)\|_{L^2(\Omega)}.$$

Hint: First assume $f \in \mathcal{D}(\Omega)$. For such f integrate $\int_0^\infty f(x)^2 x^{-2}\, dx$ by parts and apply Cauchy Schwartz to the result. Finally approximate a general f by an $\tilde{f} \in \mathcal{D}(\Omega)$.

The dual space

The dual of $C(K)$, The dual of $L^p(\Omega)$, Functionals on other spaces, The Hahn-Banach theorems, The subdifferential, Weak and weak* convergence, The weak and weak* topologies, The dual of the dual (reflexivity), The Banach-Alaoglu theorem, Application to Partial Differential Equations.

Text books to look at: Except for the last part on PDEs the material in this section is classical Functional Analysis, and is covered in great detail in [4, 5]. The part on PDEs is also classical, but traditionally does not appear in books on (linear) functional analysis.

First we identify the dual spaces of a number of Banach spaces.

40. The dual of $C(K)$

Let K be a compact metric space. If μ is a signed Borel measure on K then

$$\Lambda_\mu : f \mapsto \Lambda_\mu(f) = \int_K f(x)\, d\mu(x)$$

defines a linear functional on $C(K)$, which is bounded by

$$|\Lambda_\mu(f)| \leq \int_K |f(x)|\, d|\mu|(x) \leq |\mu|(K) \cdot \|f\|_{C(K)}.$$

We denote the space of signed Borel measures μ on K by $\mathfrak{M}(K)$.

Theorem 34. *The correspondence $\mu \in \mathfrak{M}(K) \mapsto \Lambda_\mu \in C(K)^*$ is bijective.*

I refer to Rudin's [4, theorem 6.19] for the proof.

41. The dual of $L^p(\Omega)$

For any $g \in L^q(\Omega)$ one can define a functional $\Lambda_g : L^p \to \mathbb{R}$ by

$$\Lambda_g(f) = \int_\Omega f(x)g(x)d\mu(x). \tag{12}$$

By Hölder's inequality this functional is bounded and its norm is

$$\|\Lambda_g\|_{(L^p)^*} = \|g\|_{L^q}.$$

This was proven in Lemma 10 (page 22.) Thus we have an isometric mapping of $L^q(\Omega)$ *into* the dual of $L^p(\Omega)$ for all $p \in [1, \infty]$.

Theorem 35. *For $1 \leq p < \infty$ the mapping $g \mapsto \Lambda_g$ is bijective, i.e. every bounded linear functional on $L^p(\Omega)$ is of the form Λ_g for some $g \in L^q(\Omega)$.*

Note that the case $p = \infty$ is excluded here.

Proof. I only outline a proof. For a complete proof see [4, theorem 6.16]. The proof (or better, a proof) is based on the Radon-Nikodym theorem on differentiation of measures. Given a functional $\Lambda : L^p \to \mathbb{R}$ one first defines

$$\mu(E) = \Lambda(\chi_E)$$

for any measurable $E \subset \Omega$ with $m(E) < \infty$ (m is the measure in the definition of the given L^p space.) Using the boundedness of Λ one verifies that μ is a countably additive (signed) measure on Ω. Since $\mu(E) = 0$ for any set E of vanishing m-measure the Radon-Nikodym theorem implies that $\mu(E) = \int_E g(x)dm(x)$ for some measurable $g : \Omega \to \mathbb{R}$. The proof is completed by verifying that $\Lambda(f) = \int_\Omega f(x)d\mu(x) = \int_\Omega f(x)g(x)dm(x)$. □

42. Functionals on other spaces

42.1. Sobolev spaces.

The Sobolev space $W^{1,p}(\mathbb{R}^n)$ is a subset of $L^p(\mathbb{R}^n)$ so any $g \in L^q(\mathbb{R}^n)$ with $q = \frac{p}{p-1}$ defines a linear functional on $W^{1,p}(\mathbb{R}^n)$ as in (12). However, these are not all functionals on $W^{1,p}(\mathbb{R}^n)$. First, $q = p/(p-1)$ is not the best exponent for $g \in L^q$ to define a bounded linear functional on $W^{1,p}(\mathbb{R}^n)$.

Exercise 66.
(i) Let $1 \leq p < n$, and set $r = \dfrac{np}{np + p - n}$. If $f \in L^r(\mathbb{R}^n)$ then

$$\Lambda_f(g) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} f(x)g(x)\, dx$$

defines a bounded linear functional on $W^{1,p}(\mathbb{R}^n)$.
(ii) For which values of $a > 0$ does

$$\Lambda(g) = \int_{|x|\leq 1} |x|^{-a} g(x)\, dx$$

define a bounded linear functional on $W^{1,2}(\mathbb{R}^3)$? (*Hint:* Problem (i) gives you a sufficient condition for Λ to be a bounded functional. If you think Λ is not bounded for some value of a, compute $\Lambda(\varphi_\varepsilon)$ for all $\varepsilon \in (0, 1)$, and look at how the $W^{1,p}$ norm of φ_ε depends on ε. Here $\varphi_\varepsilon = \varepsilon^{-n}\varphi(x/\varepsilon)$ is our favorite family of test functions.)

The following example shows how one can interpret "derivatives of L^q functions" as elements of the dual of $W^{1,p}(\mathbb{R}^n)$.

Exercise 67.
(i) Let $h_1, \ldots, h_n \in L^q(\mathbb{R}^n)$, where $q = \frac{p}{p-1}$. Then

$$\mathrm{M}_{h_1,\ldots,h_n}(g) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \{h_1(x)D_1g(x) + \ldots + h_n(x)D_ng(x)\}\, dx$$

defines a linear functional.
(ii) If $h_1, \ldots, h_n \in C_c^\infty$ then $\Lambda_f = \mathrm{M}_{h_1,\ldots,h_n}$, where

$$f(x) = \operatorname{div} \vec{h} = \frac{\partial h_1}{\partial x_1} + \ldots + \frac{\partial h_n}{\partial x_n}.$$

42.2. Hölder spaces.

If K is a compact metric space then $C^\alpha(K) \subset C(K)$ so that any signed measure $\mu \in \mathfrak{M}(K)$ defines a functional on $C^\alpha(K)$. Here is an example of a functional on $C^\alpha(K)$ which is not defined by a (signed) measure $\mu \in \mathfrak{M}(K)$.

Exercise 68. Let $0 < \alpha \leq 1$. For $f \in C^\alpha([-1,1])$ we define

$$\Lambda f = \lim_{\varepsilon \searrow 0} \int_{J_\varepsilon} \frac{f(x)}{x}\, dx.$$

where $J_\varepsilon = [-1, -\varepsilon) \cup (\varepsilon, 1]$.

(i) Show that Λf exists if $f \in C^\alpha([-1,1])$. (Hint: $f(x) = f(x) - f(0) + f(0)$.)

(ii) Show that Λ defines a bounded linear functional on $C^\alpha([-1,1])$.

(iii) Show that there is no $C < \infty$ such that $|\Lambda(f)| \leq C\|f\|_\infty$, and that f cannot be extended to a bounded functional on $C([-1,1])$.

42.3. Finite dimensional spaces.

There is some interesting geometry involved in finding the dual norm on finite dimensional spaces which we won't go into here, apart from the following remarks.

If $X = \mathbb{R}^n$ then any functional λ on X is completely determined by its values $\lambda_i = \lambda(e_i)$ on the standard basis. Thus we can identify X^* with $\mathbb{R}^n$, and if $\lambda = (\lambda_i)_{i=1}^n$, $x = (x_i)_{i=1}^n$, then

$$\lambda(x) = \lambda_1 x_1 + \ldots + \lambda_n x_n.$$

Exercise 69.

(i) Let $X = \mathbb{R}^3$, and let the unit ball of a given metric be the unit cube $B = [-1,1] \times [-1,1] \times [-1,1]$. Find the unit ball of the norm of the dual.

(ii) If the unit ball of a norm on $\mathbb{R}^n$ is a polyhedron given by the inequalities

$$|\lambda_1(x)| \leq 1, |\lambda_2(x)| \leq 1, \ldots, |\lambda_m(x)| \leq 1,$$

then the unit ball of the dual norm on $\mathbb{R}^n$ is the convex polyhedron with vertices λ_1, $\ldots$, λ_m. (This problem may be easier after you absorb the geometric version of the Hahn-Banach theorem, Theorem 37.)

(iii) Let X be the plane $\mathbb{R}^2$, and let the unit ball of a given norm on X be the square with vertices $(\pm a, \pm a)$ for some constant $a > 0$. Find the unit ball of the dual norm. Show that X is isometric with its dual space.

The dual of a direct sum of Banach spaces is the direct sum of their duals:

Exercise 70. If a Banach space X is the direct sum of two Banach spaces, $X = V \oplus W$, then show $X^* = V^* \oplus W^*$.

Exercise 71. What is the dual of the space of vector valued L^p functions

$$L^p(\Omega, \mathbb{R}^k) \stackrel{\text{def}}{=} \{f : \Omega \to \mathbb{R}^k \mid \int_\Omega |f(x)|\, dx < \infty\}?$$

(Hint: $L^p(\Omega, \mathbb{R}^k) = L^p(\Omega) \oplus \ldots \oplus L^p(\Omega)$.)

43. The Hahn-Banach theorems

Notice the absence of topology in the statement of the following theorem.

Theorem 36 (Hahn-Banach)**.** *Let $f : X \to \mathbb{R}$ be a convex function on a real vector space X and let $\lambda : L \to \mathbb{R}$ be a linear functional defined on some linear subspace*

$L \subset X$. If $\lambda(x) \leq f(x)$ for all $x \in L$, then there exists a linear functional $\Lambda : X \to \mathbb{R}$ with $\Lambda(x) \leq f(x)$ for all $x \in X$, and $\Lambda(x) = \lambda(x)$ for all $x \in L$.

Proof. The proof consists of two parts. First one assumes L has codimension one in X. Then one applies an induction argument on the dimension of L, or if X is infinite dimensional one applies Zorn's lemma.

We begin with the first part where we have a vector $v \in X \setminus L$, and where every $x \in X$ is of the form $x = \xi v + l$ with $l \in L$. Since $v \notin L$ this decomposition of any x is unique (why?).

The complement of L in X consists of two components,

$$X_+ \stackrel{\text{def}}{=} \{\xi v + l : l \in L, \xi > 0\}, \quad X_- \stackrel{\text{def}}{=} \{\xi v + l : l \in L, \xi < 0\}.$$

The extension Λ we are looking for is now completely determined by its value $\Lambda(v)$, since by linearity

$$\Lambda(\xi v + l) = \xi \Lambda(v) + \Lambda(l),$$

while $\Lambda(l) = \lambda(l)$ is prescribed.

We'll denote the extension for which $\Lambda(v) = c$ by Λ_c.

If we pick just any value c then the inequality $\Lambda_c(x) \leq f(x)$ may fail at certain $x \in X \setminus L$. Let

$$F_c \stackrel{\text{def}}{=} \{x \in X : \Lambda_c(x) > f(x)\}$$

denote the set where our extension fails to be bounded by f.

The set F_c is convex: if $x, y \in F_c$ then

$$\begin{aligned} f(tx + (1-t)y) &\leq tf(x) + (1-t)f(y) \\ &< t\Lambda_c(x) + (1-t)\Lambda_c(y) \\ &= \Lambda_c(tx + (1-t)y), \end{aligned}$$

so $tx + (1-t)y \in F_c$.

Since F_c is convex, and since $F_c \cap L = \varnothing$ we have either $F_c \subset X_+$ or $F_c \subset X_-$. Let $J_\pm \subset \mathbb{R}$ be the set of c for which $F_c \subset X_\pm$ and $F_c \neq \varnothing$.

The sets $J_\pm$ are disjoint. Indeed, we just observed that $F_c \subset X_+$ or $F_c \subset X_-$, but not both at the same time.

The $J_\pm$ are open. To see why J_+ is open let $c_* \in J_+$ be given, and suppose $x_* = \xi v + l \in F_{c_*}$. Then $\Lambda_c(x) = \xi c + \lambda(l)$ depends continuously on c. We have $\Lambda_{c_*}(x_*) > f(x_*)$, so for c close to c_* we will still have $\Lambda_c(x) > f(x)$: in other words, J_+ contains a neighbourhood of $c_* \in \mathbb{R}$.

The $J_\pm$ are non empty. J_+ contains the interval $(f(v), \infty)$, for if $c > f(v)$ then $\Lambda_c(v) = c > f(v)$. Likewise, J_- contains the interval $(-\infty, -f(-v))$ since $c < -f(-v)$ implies $\Lambda_c(-v) = -c > f(-v)$.

The real line is connected and hence $\mathbb{R} \neq J_+ \cup J_-$, so we have a $c_* \in \mathbb{R} \setminus J_+ \cup J_-$. By definition we have $\Lambda_{c_*}(x) \leq f(x)$ for all $x \in X$. This completes the first part of the proof.

For the second part of the proof we assume that X and L are arbitrary. Let $\mathcal{S}$ be the set of all pairs (M, Λ) of linear subspaces $L \subset M \subset X$ and linear functionals $\Lambda : M \to \mathbb{R}$ with $\Lambda \leq f$ on M. The set $\mathcal{S}$ is partially ordered by $(M, \Lambda) \lhd (M', \Lambda')$ if $M \subset M'$ and $\Lambda'|_M = \Lambda$. Every chain $\{(M_\alpha, \lambda_\alpha) : \alpha \in A\}$ has an upper bound for the ordering $\lhd$, namely $\cup_{\alpha \in A}(M_\alpha, \lambda_\alpha)$. Thus we may apply Zorn's lemma and conclude the existence of a maximal element (M, Λ) in $(\mathcal{S}, \lhd)$. For this maximal element one must have $M = X$, for otherwise one could select a $v \in X \setminus M$, define

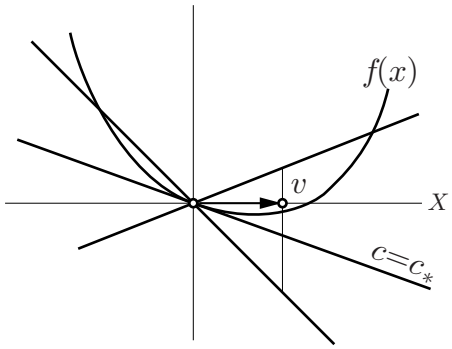


FIGURE 1. The proof when $X = \mathbb{R}$ and $L = \{0\}$.

$N = \{\xi v + m : \xi \in \mathbb{R}, m \in M\}$ and apply the first part of this proof to extend $\Lambda : M \to \mathbb{R}$ to a functional on the larger domain N, keeping $\Lambda \leq f$ all the time. This would contradict maximality of (M, Λ), so $M = X$ after all, and Λ is the extension we are after.

□

Application of the Hahn-Banach theorem to the norm $f(x) = \|x\|$ of a Banach space gives you the following alternative description of $\|x\|_X$. (Compare with Lemma 10.)

Exercise 72. Let X be a Banach space.

(i) If $x \in X$ then a functional $\lambda \in X^*$ exists with $\|\lambda\| = 1$ and $\lambda(x) = \|x\|$.

(ii) For any $x \in X$ one has $\|x\| = \max_{\|\lambda\| \leq 1} \lambda(x)$. (Note that it says "max" rather than "sup.")

(iii) Show that not every element of the dual of $L^\infty(-1, 1)$ is given by $\Lambda(g) = \int_{-1}^{1} f(x)g(x)\, dx$ for some $f \in L^1$. (Hint: use (i), (ii) and exercise 34.)

Another application of the Hahn-Banach theorem concerns finding closed complements to closed subspaces of Banach spaces. By definition, a *closed complement* of a closed subspace $L \subset X$ is another closed subspace $M \subset X$ such that $X = M \oplus L$, i.e.

$$M \cap L\{0\} \text{ and } X = M + L$$

In this case every $x \in X$ can be written in a unique manner as the sum $x = m + l$ of $m \in M$ and $l \in L$.

Exercise 73.

(i) If $L \subset X$ is closed and if $k = \dim X/L$ is finite then L has a closed complement $M \subset X$ with $\dim M = k$.

(ii) If $L \subset X$ is finite dimensional then L has a closed complement M. (Hint: use Hahn-Banach and write $M = \{x : \lambda_1(x) = \ldots = \lambda_k(x) = 0\}$ for certain $\lambda_1, \ldots, \lambda_k \in X^*$, where $k = \dim L$.)

The following is sometimes called the geometric version of the Hahn-Banach theorem.

Theorem 37. *Let C be a closed and convex subset of the Banach space X, and let $x_0 \in X \setminus C$. Then there exist $\lambda \in X^*$ and $t \in \mathbb{R}$ with*

$$\lambda(x) < t < \lambda(x_0)$$

for all $x \in C$. In other words, the hyperplane $\{x \in X : \lambda(x) = t\}$ separates C and x_0.

Proof. The statement is invariant under translations in X so we may assume that the origin belongs to C.

Since C is closed an $\varepsilon > 0$ exists with $B(x_0, 2\varepsilon) \cap C = \varnothing$. Consider the set

$$C_\varepsilon = \bigcup_{x \in C} B(x, \varepsilon).$$

Then C_ε is open and convex, and $B(x_0, \varepsilon) \cap C = \varnothing$.

Define $f : X \to \mathbb{R}$ by

$$f(x) = \inf\{r > 0 : x \in rC_\varepsilon\}.$$

This function is convex, and since $B(0, \varepsilon) \subset C_\varepsilon$, it satisfies

$$f(x) \leq \frac{1}{\varepsilon}\|x\|_X. \tag{13}$$

It is also almost homogeneous in that $f(tx) = tf(x)$ for all $t > 0$ (but not necessarily for $t < 0$ since we don't know if C_ε is symmetric; see §15.)

On the one-dimensional subspace $L \subset X$ spanned by x_0 we define a linear functional λ by setting $\lambda(x_0) = f(x_0)$. One then has $\lambda \leq f$ on L, and by Hahn-Banach an extension $\Lambda : X \to \mathbb{R}$ exists with $\Lambda \leq f$. By (13) we also have

$$\Lambda(x) \leq \frac{1}{\varepsilon}\|x\|$$

so that Λ is bounded.

To complete the proof we note that $C \subset C_\varepsilon$, and that on C_ε one has $f(x) \leq 1$. On the other hand one has $f(x_0) > 1$, so

$$\sup_{x \in C} \Lambda(x) \leq 1 < \Lambda(x_0).$$

□

Exercise 74.

(i) If $V \subset \mathbb{R}^3$ is a plane through the origin then the intersection of V with the unit cube $B = [-1, 1]^3$ is a polygon. How many sides can this polygon have?

(ii) Let $X = \mathbb{R}^n$ with the maximum norm $\|(x_i)\| = \max_{1 \leq i \leq n} |x_i|$. Denote the unit ball in X by B. If $T : \mathbb{R}^2 \to X$ is a linear map, then $T^{-1}(B)$ is a convex polygon. Which polygons are of the form $T^{-1}(B)$?

(iii) Suppose $T : \mathbb{R}^2 \to \ell^\infty$ is a linear map. Then $T^{-1}(B)$ is a closed convex subset of $\mathbb{R}^2$ (B denotes the unit ball in ℓ^∞.) Which closed convex subset of $\mathbb{R}^2$ are of the form $T^{-1}(B)$?

(iv)* Let X be a separable Banach space. Show that there exists an isometric embedding $T : X \hookrightarrow \ell^\infty$.

Besides the extensions of linear functionals provided by the Hahn-Banach theorem there is a simpler extension theorem which is often useful:

Exercise 75. Let $D \subset X$ be a dense linear subspace of the Banach space X, and let $\lambda : D \to \mathbb{R}$ be a linear functional which satisfies $|\lambda(x)| \leq M\|x\|$ for all $x \in D$. Show that there is a unique bounded linear functional $\Lambda \in X^*$ with $\Lambda(x) = \lambda(x)$ for all $x \in D$.

44. The dual of $W^{1,p}(\Omega)$

In exercise 67 we saw that for any $h_0, \ldots, h_n \in L^q(\mathbb{R}^n)$, with $q = \frac{p}{p-1}$ the expression

$$\mathrm{M}_{h_0,\ldots,h_n}(g) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} \{h_0(x)g(x) + h_1(x)D_1g(x) + \ldots + h_n(x)D_ng(x)\}\, dx$$

defines a bounded linear functional on $W^{1,p}(\Omega)$, and that the operator norm of $\mathrm{M}_{h_0,\ldots,h_n}$ is bounded by

$$\|\mathrm{M}_{h_0,\ldots,h_n}\|_{(W^{1,p})^*} \leq C \sum \|h_i\|_{L^q}.$$

Using the Hahn-Banach theorem we can show that there are no other bounded linear functionals on $W^{1,p}(\Omega)$.

Theorem 38. *Every* $\mathrm{M} \in (W^{1,p}(\Omega))^*$ *is of the form* $\mathrm{M} = \mathrm{M}_{h_0,\ldots,h_n}$ *for certain* $h_i \in L^{p/(p-1)}(\Omega)$.

Proof. We begin by showing that $W^{1,p}(\Omega)$ is a closed subspace of

$$X = \overbrace{L^p(\Omega) \oplus \ldots \oplus L^p(\Omega)}^{n+1 \text{ terms}}.$$

Namely, one assigns to any function $f \in W^{1,p}(\Omega)$ the $n+1$ tuple $(f, D_1f, \ldots, D_nf) \in X$. Not every $n+1$ tuple $(f, g_1, \ldots, g_n) \in X$ can be realized in this way. In fact the definition of distributional derivative says that $(f, g_1, \ldots, g_n) = (f, D_1f, \ldots, D_nf)$ if and only if $\int_\Omega fD_i\varphi - g_i\varphi = 0$ holds for all test functions $\varphi \in \mathcal{D}(\Omega)$. Thus one can identify $W^{1,p}(\Omega)$ with

$$Y = \left\{ (f, g_1, \ldots, g_n) \in X : \forall_{\varphi \in \mathcal{D}(\Omega)} \int_\Omega fD_i\varphi - g_i\varphi = 0 \right\}.$$

This subspace of X is closed, as you can check for yourself.

If M is a bounded linear functional on $W^{1,p}(\Omega)$ then we can regard it as a bounded linear functional on Y, and by the Hahn-Banach theorem we may assume that M is the restriction of some bounded linear functional on X, whcih we again denote by M.

The proof is completed by observing that bounded linear functional on X are all of the form

$$\mathrm{M}(g_0, \ldots, g_n) = \int_{\mathbb{R}^n} \{h_0(x)g_0(x) + h_1(x)g_1(x) + \ldots + h_n(x)g_n(x)\}\, dx.$$

for certain $h_i \in L^q(\Omega)$ □

Exercise 76. Show that for $1 < p < \infty$ the dual of $W^{1,p}(\Omega)$ is a separable Banach space.

45. The subdifferential

If $\Phi : X \to \mathbb{R}$ is a convex function on some Banach space then the *subdifferential* of Φ at some $x_0 \in X$ is the collection of all $\lambda \in X^*$ such that

$$\Phi(x) \geq \Phi(x_0) + \lambda(x - x_0) \tag{14}$$

holds for all $x \in X$. Notation:

$$\partial\Phi(x_0) \stackrel{\text{def}}{=} \{\lambda \in X^* : (14) \text{ holds for all } x \in X^*\}.$$

Theorem 39. *If $\Phi : X \to \mathbb{R}$ is a continuous and convex function, then the subdifferential $\partial\Phi(x_0)$ is nonempty for all $x_0 \in X$.*

Proof. Given x_0 we consider the function $\Psi(x) = \Phi(x_0 + x) - \Phi(x_0)$. This function is also convex and continuous.

The Hahn-Banach theorem gives us a linear functional $\lambda : X \to \mathbb{R}$ which satisfies $\lambda(x) \leq \Psi(x)$ for all $x \in X$. Consequently (14) holds for λ. To complete the proof we must show that the linear functional λ is bounded (Hahn-Banach does not guarantee this).

Since Ψ is continuous there is a $\delta > 0$ such that $\Psi(x) \leq 1$ for all x with $\|x\| < \delta$. Hence $\lambda(x) < 1$ for $\|x\| < \delta$, and therefore, by homogeneity of λ and the norm on X, $|\lambda(x)| \leq \delta^{-1}\|x\|$ for arbitrary $x \in X$. □

Exercise 77.
(i) If $\Phi : \mathbb{R} \to \mathbb{R}$ is convex, and if Φ is differentiable at $x_0 \in \mathbb{R}$ then $\partial\Phi(x_0) = \{\Phi'(x_0)\}$.
(ii) Compute $\partial\Phi(x)$ if $\Phi : \mathbb{R} \to \mathbb{R}$ is given by $\Phi(x) = |x|$.
(iii) Compute $\partial\Phi(x_0)$ at $x_0 = (1, 1)$ if $\Phi : \mathbb{R}^2 \to \mathbb{R}$ is given by $\Phi(x, y) = \max\{|x|, |y|\}$ (i.e. Φ is the ℓ^∞ norm on $\mathbb{R}^2$.)

46. Weak and weak* convergence

Let X be a Banach space.

A sequence $\{x_n \in X : n \in \mathbb{N}\}$ *converges weakly to $x \in X$,* written $x_n \rightharpoonup x$, if for every $\lambda \in X^*$ one has $\lim_{n\to\infty} \lambda(x_n) = \lambda(x)$.

A sequence of functionals $\lambda_n \in X^*$ *converges weak* to $\lambda \in X^*$,* written $\lambda_n \overset{*}{\rightharpoonup} \lambda$, if for any $x \in X$ one has $\lim_{n\to\infty} \lambda_n(x) = \lambda(x)$.

Exercise 78. Show that any norm convergent sequence $\{x_i \in X : i \in \mathbb{N}\}$ is weakly convergent. Show that any norm convergent sequence $\{\lambda_i \in X^* : i \in \mathbb{N}\}$ is weak* convergent.

Exercise 79.
(i) Show that the sequence $f_k(x) = \chi_{[k,k+1]}(x)$ converges weakly to zero in $L^p(\mathbb{R})$ if $1 < p < \infty$. What happens when $p = 1$?
(ii) Does the sequence $f_k(x) = k\chi_{[k,k+1]}(x)$ converge weakly to zero in $L^p(\mathbb{R})$?
(iii) Does the sequence $f_k(x) = (\ln\ln k)\chi_{[k,k+1]}(x)$ converge weakly to zero in $L^p(\mathbb{R})$?

We have seen that $L^p(\Omega)$ is also the dual space of $L^q(\Omega)$ if $q = p/(p-1)$, $1 < p \leq \infty$. We can therefore speak of weak and of weak* convergence of sequences in $L^p(\Omega)$.

Exercise 80. Show that if $1 < p < \infty$ the notions of weak and of weak* convergence coincide on $L^p(\Omega)$.

The following exercises show that the notions of weak convergence in L^p and convergence in the sense of distributions are very similar, but not the same.

Exercise 81. Let $\Omega \subset \mathbb{R}^n$ be an open domain, and let $1 \leq p < \infty$. Let $f_k \in L^p(\Omega)$ be a bounded sequence. Show that the sequence f_k converges weakly in L^p to $f \in L^p$ if and only if the f_k converge in the sense of distributions to $f \in L^p$

Exercise 82.
(i) (The Riemann-Lebesgue Lemma) We have seen that one can identify $L^\infty(\mathbb{R})$ with the dual space of $L^1(\mathbb{R})$. Show that the sequence $f_n(x) = \sin nx$ converges weak* to zero in L^∞.
(ii) Show that the sequence $g_n(x) = n \sin nx$ does NOT converge weak* to zero.

Exercise 83. Let $1 < p < \infty$.

Suppose a sequence of functions $f_k \in L^p(\mathbb{R}^n)$ converges in the sense of distributions to $T \in \mathcal{D}'(\mathbb{R}^n)$. Suppose also that the sequence f_k is bounded. Show that the limit distribution T can actually be represented by an $f \in L^p(\mathbb{R}^n)$.

The following is a more abstract version of the previous exercise:

Exercise 84. Let D be a dense subset of a Banach space X, and let $\{\lambda_i : i \in \mathbb{N}\}$ be a sequence of linear functionals which (i) are uniformly bounded:

$$\sup \|\lambda_i\|_{X^*} = M < \infty$$

and (ii) converge pointwise on D, i.e.

$$\Lambda(x) = \lim_{i\to\infty} \lambda_i(x) \text{ exists for all } x \in D.$$

Show that this limit exists for all $x \in X$, that the limit Λ is a bounded linear functional, and that $\lambda_i \overset{*}{\rightharpoonup} \Lambda$.

Exercise 85. *(i)* Let $X = C([-1,1])$ and consider the linear functionals $\Lambda_n, \Lambda \in X^*$ defined by

$$\Lambda_n(f) \overset{\text{def}}{=} f(1/n), \qquad \Lambda(f) \overset{\text{def}}{=} f(0).$$

Show that $\Lambda_n \overset{*}{\rightharpoonup} \Lambda$ as $n \nearrow \infty$, but that Λ_n does not converge in the norm of X^* to Λ.

(ii) Let Y be the Hölder space $C^\alpha([-1,1])$. The Λ_n and Λ as defined in (i) also define bounded linear functionals on Y. Show that Λ_n converges to Λ in the norm of Y^*, i.e. show

$$\lim_{n\to\infty} \|\Lambda_n - \Lambda\|_{C^\alpha([-1,1])^*} = 0.$$

47. The weak and weak* topologies.

The usual approach to defining convergence of sequences in a set X involves the introduction of a topology on X. One can indeed introduce topologies on X and X^* such that weak and weak* convergence are simply convergence with respect to these topologies.

The weak topology on X is defined by specifying that $U \subset X$ is a neighbourhood of 0 if there exist functionals $\lambda_1, \ldots, \lambda_m \in X^*$ and an $\varepsilon > 0$ such that

$$V(\lambda_1, \ldots, \lambda_m, \varepsilon) \overset{\text{def}}{=} \{x \in X : |\lambda_1(x)| < \varepsilon, \ldots, |\lambda_m(x)| < \varepsilon\}$$

is contained in U. A set $U \subset X$ is a neighborhood of some $x_0 \in X$ if $U - x_0 = \{x - x_0 : x \in X\}$ is a neighborhood of 0.

The weak* topology is defined in the same way: a $U \subset X^*$ is a neighbourhood of 0 if there exist $x_1, \ldots, x_m \in X$ and an $\varepsilon > 0$ such that

$$W(x_1, \ldots, x_m, \varepsilon) \overset{\text{def}}{=} \{\lambda \in X^* : |\lambda(x_1)| < \varepsilon, \ldots, |\lambda(x_m)| < \varepsilon\}$$

is contained in U. A set $U \subset X^*$ is a neighborhood of some $\lambda_0 \in X$ if $U - \lambda_0 = \{\lambda - \lambda_0 : \lambda \in X^*\}$ is a neighborhood of 0.

The vector spaces X and X^* with their weak and weak* topologies respectively are examples of locally convex topological vector spaces (LCTVS). For a careful treatment of their theory you should look at Rudin's [5].

For general Banach spaces X the weak topology need not be metrizable, so that one cannot test for openness or closedness of subsets $A \subset X$ by looking at convergent sequences only.

We will not use the theory of LCTVS's, and will only use the following observation.

Exercise 86.

(i) Let $\lambda \in X^*$ and let $a \in \mathbb{R}$ be a constant. Show that the set $\{x \in X : \lambda(x) < a\}$ is an open subset of X in the weak topology. Show that $\{x \in X : \lambda(x) \leq a\}$ is closed in the weak topology.

(ii) Let $x \in X$ and let $a \in \mathbb{R}$ be a constant. Show that the set $\{\lambda \in X^* : \lambda(x) < a\}$ is an open subset of X^* in the weak* topology. Show that $\{\lambda \in X^* : \lambda(x) \leq a\}$ is closed in the weak* topology.

The following is an application of the geometric version of the Hahn-Banach theorem and the previous exercise.

Exercise 87. Let $C \subset X$ be a convex subset of a Banach space which is closed with respect to the norm topology. Show that C is also closed in the weak topology.

Show that any linear subspace $L \subset X$ is closed in the norm topology if and only if it is closed in the weak topology.

48. The dual of the dual

If X is a Banach space then any $x \in X$ defines a linear functional $\hat{x}$ on X^* by

$$\hat{x}(\lambda) = \lambda(x).$$

This functional is bounded by $|\hat{x}(\lambda)| \leq \|\lambda\|_{X^*}\|x\|_X$, so that $\hat{x}$ belongs to the dual of X^*, and $\|\hat{x}\|_{X^{**}} \leq \|x\|_X$. Thus the map $x \mapsto \hat{x}$ defines a bounded linear map from X into X^{**}. This map is in fact an isometry since for each $x \in X$ a $\lambda \in X^*$ can be found with $\lambda(x) = \|x\|$ and $\|\lambda\|_{X^*} = 1$, so that $\|\hat{x}\| \geq \hat{x}(\lambda) = \lambda(x) = \|x\|$.

By definition a Banach space is called *reflexive* if the embedding $X \hookrightarrow X^{**}$ is surjective, i.e. if every bounded linear functional on X^* is of the form $\hat{x}$ for some $x \in X$.

Exercise 88.

(i) Show that $L^p(\Omega)$ is reflexive if $1 < p < \infty$, but not when $p = 1$.

(ii) Show that $C(K)$ is not reflexive.

Exercise 89.

(i) Show that a closed subspace $L \subset X$ of a reflexive Banach space is also reflexive.

(ii) Show that if X and Y are reflexive Banach spaces then the sum $X \oplus Y$ is also reflexive.

(iii) Show that $W^{1,p}(\Omega)$ is reflexive if $1 < p < \infty$.

49. The Banach-Alaoglu theorem

Theorem 40 (Banach-Alaoglu)**.** *The unit ball of the dual of any Banach space is compact in the weak* topology.*

The proof can be found in [5]. The proof relies on Tychonov's theorem which says that the product $\prod_{\alpha \in \mathrm{A}} X_\alpha$ of a family of compact topological spaces X_α is again compact in the product topology (with no restrictions on how many there are: A could be an uncountably infinite set!) This theorem from point-set topology in turn relies on Zorn's Lemma (or the Axiom of Choice, or Hausdorff's maximality principle.) Since I'm avoiding point-set topology in this class we will prove the following "sequential compactness version" of this theorem.

Theorem 41. *Let X be a separable Banach space. Any bounded sequence of functionals $\{\lambda_n \in X^* : n \in \mathbb{N}\}$ has a weak* convergent subsequence.*

Proof. Let $\{x_k : k \in \mathbb{N}\}$ be a dense sequence in X, and let λ_i be the given sequence of functionals. By assumption they are bounded, so we have $\|\lambda_i\|_{X^*} \leq M$ for some $M < \infty$.

We can extract a subsequence λ_{i_j} such that $\lim_{j\to\infty} \lambda_{i_j}(x_1)$ exists. From this subsequence we can extract a further subsequence such that $\lambda_{i_j^{(2)}}(x_1)$ and $\lambda_{i_j^{(2)}}(x_2)$ both converge. Proceeding by induction we obtain a sequence of subsequences $\lambda_{i_j^{(l)}}$ for which

$$\lim_{j\to\infty} \lambda_{i_j^{(l)}}(x_m)$$

exists if $m \leq l$. Cantor's diagonal trick then gives us a subsequence $\lambda_{i_j''}$ of λ_i for which

$$\Lambda(x_k) \stackrel{\text{def}}{=} \lim_{j\to\infty} \lambda_{i_j''}(x_k)$$

exists for all $k \in \mathbb{N}$.

We now show that $\lambda_{i_j''}(x)$ converges for all $x \in X$.

Let $\varepsilon > 0$ and $x \in X$ be given. The x_k are dense in X so we can find an x_k with $\|x - x_k\| < \epsilon/4M$. Since $\lambda_{i_j''}(x_k)$ converges as $j \nearrow \infty$ we can find $N < \infty$ such that

$$|\lambda_{i_l''}(x_k) - \lambda_{i_m''}(x_k)| < \varepsilon/2$$

for all $l, m > N$. The uniform boundedness of the λ_j now implies that for all $l, m > N$ one has

$$\begin{aligned} |\lambda_{i_l''}(x) - \lambda_{i_m''}(x)| &\leq |\lambda_{i_l''}(x) - \lambda_{i_l''}(x_k)| + |\lambda_{i_l''}(x_k) - \lambda_{i_m''}(x_k)| \\ &\quad + |\lambda_{i_m''}(x_k) - \lambda_{i_m''}(x)| \\ &< M\frac{\varepsilon}{4M} + \frac{\varepsilon}{2} + M\frac{\varepsilon}{4M} \\ &= \varepsilon. \end{aligned}$$

Thus $\lambda_{i_j''}(x)$ is a Cauchy sequence and

$$\Lambda(x) \stackrel{\text{def}}{=} \lim_{j\to\infty} \lambda_{i_j''}(x)$$

exists for all $x \in X$.

Uniform boundedness of the λ_i implies

$$|\Lambda(x)| \leq \sup_{j\in\mathbb{N}} |\lambda_{i_j''}(x)| \leq M\|x\|$$

so that Λ is a bounded functional on X. By definition $\lambda_{i_l''}$ converges weak* to Λ. □

The "dual theorem" to the Banach-Alaoglu theorem would state that the unit ball in any Banach space is weakly compact, or, any bounded sequence in a Banach space would have a weakly convergent subsequence. This is not true in general, but we can prove the following

Theorem 42. *If X is a reflexive Banach space whose dual is separable, then any bounded sequence in X has a weakly convergent subsequence.*

Proof. The trick is to identify X with X^{**}. If $x_i \in X$ is a bounded sequence, then we consider the functionals $\hat{x}_i \in X^{**}$ defined by $\hat{x}_i(\lambda) = \lambda(x_i)$ for all $\lambda \in X^*$.

The Banach-Alaoglu theorem gives us a subsquence $\hat{x}_{i_j}$ which converges weak* to some $\xi_\infty \in X^{**}$. But since X is reflexive there is an $x_\infty \in X$ such that $\xi_\infty = \hat{x}_\infty$. Weak* convergence in X^{**} of the $\hat{x}_{i_j}$ to ξ_∞ then implies weak convergence in X of the x_{i_j} to x_∞. □

The following example shows that things do indeed go wrong if the Banach space is not reflexive.

Exercise 90. Show that the sequence $f_n(x) = n\chi_{(0,1/n)}(x)$ is bounded in $L^1(\mathbb{R})$ but does not have a weakly convergent subsequence. (Hint: what would be the support of the weak limit?)

An attempt to construct a similar example in $L^p(\mathbb{R})$ leads to the following problem:

Exercise 91. Prove that the sequence $f_n(x) = n^{1/p}\chi_{(0,1/n)}(x)$ converges weakly to 0 in $L^p(\mathbb{R})$.

50. Application to Partial Differential Equations

50.1. A general minimization theorem.

We return to the problem of finding the minimum of a continuous function f on a subset K of some Banach space. In general, even if $\inf_K f(x) > -\infty$ the function f need not attain its minimum, even for fairly simple f and K (see exercise 34 again). Using the Banach-Alaoglu theorem one can give sufficient conditions for a function to attain its minimum.

Theorem 43. *Let X be a reflexive Banach space whose dual is separable. Let $\Phi : X \to \mathbb{R}$ be a continuous and convex function. Assume furthermore that Φ satisfies*

$$\lim_{\substack{\|x\|\to\infty \\ x\in K}} \Phi(x) = \infty. \tag{15}$$

Then $\inf_{x\in X} \Phi(x) > -\infty$. Moreover, for any closed and convex subset $K \subset X$ an $x_0 \in K$ exists which minimizes Φ on K, i.e.

$$\Phi(x_0) = \inf_{x\in K} \Phi(x).$$

If Φ is strictly convex on K then the minimizer x_0 is unique.

Condition (15) is sometimes called "coercivity" of Φ on K.

By definition Φ is *strictly convex on* K if for any $x \neq y$ in K one has

$$\Phi(tx + (1-t)y) < t\Phi(x) + (1-t)\Phi(y) \text{ for all } 0 < t < 1. \tag{16}$$

Proof. If Φ is strictly convex then Φ cannot have two distinct minimizers in K, for if both x and y were minimizers with $x \neq y$ then (16) would imply that $\Phi(z) < \Phi(x)$ for $z = \frac{1}{2}(x+y)$, in contradiction with the minimizing property of x.

We now worry about the existence of a minimizer.

To show that Φ is bounded from below we choose $R < \infty$ so large that $\Phi(x) \geq 0$ for $\|x\| \geq R$. By theorem 39 the subdifferential of Φ is never empty so we can choose a $\lambda \in \partial\Phi(0)$. Then

$$\Phi(x) \geq \Phi(0) + \lambda(x) \geq \Phi(0) - R\|\lambda\|$$

whenever $\|x\| \leq R$. Thus Φ is bounded from below, and for any given $K \subset X$

$$\Phi_{\min} \stackrel{\text{def}}{=} \inf_{x \in K} \Phi(x) \geq \Phi(0) - R\|\lambda\|$$

is well defined.

To show that the infimum is attained we let $x_i \in K$ be a sequence with $\Phi(x_i) \to \Phi_{\min}$. Such a sequence must be bounded since our function Φ is by assumption coercive, i.e. (15) implies that there is an $R < \infty$ such that $\Phi(x) \geq \Phi_{\min} + 1$ for all x with $\|x\| \geq R$. With a finite number of exceptions all x_i therefore satisfy $\|x_i\| \leq R$.

By Banach-Alaoglu the sequence has a weakly convergent subsequence $x_{i_j} \rightharpoonup x_{\min} \in X$. Since K is closed and convex the Hahn-Banach theorem implies that the weak limit $x_{\min}$ lies in K. We will show that $x_{\min}$ is a minimizer for Φ.

Choose any $\lambda \in \partial\Phi(x_{\min})$. Then one has

$$\begin{aligned}
\Phi_{\min} &= \lim_{j \to \infty} \Phi(x_{i_j}) \\
&\geq \lim_{j \to \infty} \Phi(x_{\min}) + \lambda(x_{i_j} - x_{\min}) \\
&= \Phi(x_{\min}) \\
&\geq \Phi_{\min}.
\end{aligned}$$

So we do indeed have $\Phi(x_{\min}) = \Phi_{\min}$. □

50.2. A modified Dirichlet problem.

One can use the theory we have developed so far to prove existence of solutions to certain boundary value problems. To illustrate this we consider the following problem: Given a domain $\Omega \subset \mathbb{R}^n$ and a function $g : \partial\Omega \to \mathbb{R}$ find a function $u : \Omega \to \mathbb{R}$ which satisfies

$$\begin{cases} \Delta u - u = 0 \\ u\,|_{\partial\Omega} = g \end{cases} \tag{17}$$

To be more precise we will look for solutions $u \in W^{1,2}(\Omega)$ of $\Delta u - u = 0$, i.e. we will look for functions $u \in W^{1,2}(\Omega)$ which satisfy $\Delta u - u = 0$ in the sense of distributions. To specify the boundary conditions we assume that not g, but a function $G \in W^{1,2}(\Omega)$ is given instead, and we will require that $u - G \in W^{1,2}_o(\Omega)$.

We write

$$L_G \stackrel{\text{def}}{=} \{u \in W^{1,2}(\Omega) : u - G \in W^{1,2}_o(\Omega)\},$$

i.e. L_G is the set of $u \in W^{1,2}(\Omega)$ which satisfy the boundary condition "$u = g$."

Lemma 44. *A function $u \in L_G$ satisfies* (17) *if and only if it minimizes the quantity*

$$Q(u) \stackrel{\text{def}}{=} \tfrac{1}{2} \int_\Omega \{u(x)^2 + |\nabla u(x)|^2\}\, dx$$

among all $u \in L_G$.

The proof goes exactly as the proof of Theorem 3 of section 8.

To solve the boundary value problem we must therefore find a $u \in L_G$ which minimizes Q. But Q is the square of the norm on $W^{1,2}_o$ and hence is convex and continuous. The subset L_G is closed and convex. We can therefore apply Theorem 43 with $\Phi = Q$ and $K = L_G$ and immediately conclude that a minimizer $u \in L_G$

exists, so that we have shown that *for all possible boundary data G the boundary value problem* (17) *has a solution* $u \in W^{1,2}(\Omega)$.

To conclude we observe that Q is strictly convex. Indeed, for $u \neq v \in W^{1,2}(\Omega)$ one has

$$q(t) \stackrel{\text{def}}{=} Q(tu + (1-t)v) = Q(v + t(u-v)) = A + Bt + Ct^2$$

where $A = Q(v)$ and B don't really matter and

$$C = \frac{1}{2}\int_\Omega \{(u-v)^2 + |\nabla(u-v)|^2\}\, dx > 0$$

since $u \neq v$. The quadratic function $q(t)$ is therefore strictly convex, and hence $Q(tu + (1-t)v) < tQ(u) + (1-t)Q(v)$ for $0 < t < 1$.

The minimizer $u \in L_G$ is therefore unique and we conclude that for all possible boundary data G the boundary value problem (17) has *exactly one* solution $u \in W^{1,2}(\Omega)$.

50.3. The Dirichlet Problem.

We now show how one can find harmonic functions with prescribed boundary values by proving that the Dirichlet functional attains its minimum. We apply the same minimization theorem as before to the Dirichlet functional

$$D(u) = \tfrac{1}{2}\int_\Omega |\nabla u(x)|^2\, dx.$$

This functional is almost the same as the functional Q we have considered in the previous section, the difference being the u^2 term in the integral. The absence of this term makes that the Dirichlet functional does not satisfy the "coercivity" condition (15) on arbitrary domains Ω. However, if the domain is bounded then we can still prove (15) for D. The key is *Poincaré's inequality:*

Lemma 45 (Poincaré)**.** *Let Ω be a connected open subset of $\mathbb{R}^n$ with $|\Omega| < \infty$. Then any $u \in W_o^{1,2}(\Omega)$ satisfies*

$$\int u(x)^2\, dx \leq C \int |\nabla u(x)|^2\, dx. \tag{18}$$

for some C which does not depend on u.

Proof. Since $\mathcal{D}(\Omega)$ is dense in $W_o^{1,2}(\Omega)$ we may assume that $u \in \mathcal{D}(\Omega)$.

When $n = 1$ we may assume that $\Omega = (0, L)$. For $u \in \mathcal{D}(\Omega)$ one then has

$$u(x)^2 = \int_0^x 2u(\xi)u'(\xi)\, d\xi \leq 2\|u\|_{L^2(\Omega)}\|u'\|_{L^2(\Omega)}.$$

Integrate over $x \in (0, L)$ to get

$$\|u\|^2_{L^2(\Omega)} \leq 2L\|u\|_{L^2(\Omega)}\|u'\|_{L^2(\Omega)},$$

which implies

$$\|u\|_{L^2(\Omega)} \leq 2L\|u'\|_{L^2(\Omega)}.$$

When $n \geq 2$ Hölder's inequality implies that

$$\int_\Omega |\nabla u(x)|^{\frac{2n}{n+2}}\, dx \leq \left(\int_\Omega |\nabla u(x)|^2\, dx\right)^{\frac{n}{n+2}} |\Omega|^{\frac{2}{n+2}} \tag{19}$$

The Sobolev inequality (Theorem 27) with $r = 2$ and $p = \frac{2n}{n+2}$ implies

$$\|u\|_{L^2} \leq 2(1 - \tfrac{1}{n})\|\nabla u\|_{L^{2n/(n+2)}}$$

which, combined with (19) gives

$$\|u\|_{L^2} \leq 2(1 - \tfrac{1}{n})|\Omega|^{1/n}\|\nabla u\|_{L^2}.$$

□

With this lemma we can attack the Dirichlet problem, which we formulate as follows: Let Ω be a bounded open domain in $\mathbb{R}^n$, and let $G \in W_o^{1,2}(\Omega)$ be given. Find a function $u \in W^{1,2}(\Omega)$ with

$$u - G \in W_o^{1,2}(\Omega) \text{ (boundary condition)}$$

and

$$\Delta u = 0 \text{ (Laplace's equation)}$$

in the sense of distributions.

As we observed in section 8, solutions of this problem are exactly the functions $u \in L_G$ which minimize $D(u)$.

Theorem 46. *For any $G \in W_o^{1,2}(\Omega)$ there is a unique minimizer $u_G \in L_G$ of $D(u)$.*

Proof. Uniqueness follows from the strict convexity of the Dirichlet functional.

Convexity and continuity of $D : W^{1,2}(\Omega) \to \mathbb{R}$ are proved as before.

To establish the coercivity condition we use Poincaré's inequality to estimate the $W^{1,2}$ norm of any $u \in L_G$ in terms of $D(u)$.

$$\begin{aligned}
\|u\|_{W^{1,2}}^2 &= \|u\|_{L^2}^2 + \underbrace{\|\nabla u\|_{L^2}}_{=2D(u)} \\
&= \|u - G + G\|_{L^2}^2 + 2D(u) \\
&\leq \underbrace{2\|u - G\|_{L^2}^2}_{\text{apply Poincaré}} + 2\|G\|_{L^2}^2 + 2D(u) \\
&\leq C\|\nabla(u - G)\|_{L^2} + 2\|G\|_{L^2}^2 + 2D(u) \\
&\leq 2C\|\nabla u\|_{L^2} + 2C\|\nabla G\|_{L^2} + 2\|G\|_{L^2}^2 + 2D(u) \\
&\leq C'\|G\|_{W^{1,2}} + C''D(u).
\end{aligned}$$

It follows that $\|u\|_{W^{1,2}} \to \infty$ will force $D(u) \to \infty$. We may therefore apply Theorem 43 and conclude the existence of a minimizer $u \in L_G$ for $D(u)$. □

Exercise 92. Check that D is indeed strictly convex.

50.4. A third example – Poisson's equation

Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. For any given $f : \Omega \to \mathbb{R}$ we try to find a function $u : \Omega \to \mathbb{R}$ which satisfies

$$\begin{cases} -\Delta u = f \text{ in } \Omega, \\ u\,|_{\partial\Omega} = 0. \end{cases} \tag{20}$$

This equation is called Poisson's equation. One interpretation of this equation is from electrostatics: If a cavity Ω contains a charge distribution $f(x)$, and the boundary $\partial\Omega$ of the cavity is "grounded," i.e. its potential u is kept at 0 (Volts), then the electric potential in the cavity generated by the charge distribution is precisely the solution $u(x)$ of (20).

To solve the problem we observe

Lemma 47. *Let $f \in L^2(\Omega)$. A function $u \in W_o^{1,2}(\Omega)$ which minimizes*

$$Q_f(u) = \int_\Omega \{\tfrac{1}{2}|\nabla u|^2 - f(x)u(x)\}\, dx$$

over all $u \in W_o^{1,2}(\Omega)$ satisfies $-\Delta u = f$ in the sense of distributions.

Proof. Let $\varphi \in \mathcal{D}(\Omega)$ be arbitrary and expand

$$\left(\frac{dQ_f(u+t\varphi)}{dt}\right)_{t=0} = 0.$$

One finds

$$\int_\Omega \{\nabla u \cdot \nabla\varphi - f(x)\varphi(x)\}\, dx = 0,$$

i.e.

$$\langle -\Delta u - f, \varphi\rangle = 0$$

as claimed. □

Lemma 48. *Q_f is continuous, strictly convex and coercive on $X = W_o^{1,2}(\Omega)$*

Proof. The first term in Q_f is just the Dirichlet integral, while the last term is a linear functional on $W_o^{1,2}(\Omega)$. Thus we can write Q_f as

$$Q_f(u) = D(u) + \varphi(u),$$

where

$$\varphi(u) = \int_\Omega f(x)u(x)\, dx.$$

The Dirichlet integral is a strictly convex function and linear functionals are convex so that Q_f is clearly convex.

We have alreasy verified continuity of the Dirichlet functional. The linear functional φ is bounded by

$$\left|\int_\Omega f(x)u(x)\, dx\right| \le \|f\|_{L^2}\|u\|_{L^2} \le \|f\|_{L^2}\|u\|_{W^{1,2}}$$

and is therefore also continuous. So Q_f is continuous.

To prove coercivity we recall that Poincaré's inequality implies

$$\|u\|^2_{W_o^{1,2}} = \int_\Omega \{|u|^2 + |\nabla u|^2\}\, dx \le \left(1 + 2|\Omega|^{2/n}\right)\int |\nabla u|^2\, dx$$

i.e.

$$\|u\|^2_{W_o^{1,2}} \le \left(1 + 2|\Omega|^{2/n}\right)D(u).$$

Coercivity of Q_f then follows from

$$\begin{aligned} Q_f(u) &= D(u) + \int f u \, dx \\ &\geq c\|u\|^2_{W^{1,2}} - \|f\|_{L^2}\|u\|_{W^{1,2}} \qquad \text{use } 2ab \leq \epsilon a^2 + \frac{1}{\epsilon}b^2 \\ &\geq c\|u\|^2_{W^{1,2}} - \frac{c}{2}\|u\|^2_{L^2} - \frac{1}{2c}\|f\|^2_{L^2} \\ &\geq \frac{c}{2}\|u\|^2_{W^{1,2}} - \frac{1}{2c}\|f\|^2_{L^2}. \end{aligned}$$

□

We now apply theorem 43 to $\Phi = Q_f$ and $K = X = W^{1,2}_o(\Omega)$ and immediately conclude that (20) has a solution $u \in W^{1,2}_o(\Omega)$ for any $f \in L^2(\Omega)$.

Exercise 93. Let $g_1, \ldots, g_n \in L^2(\Omega)$ be given functions. Show that a solution u of

$$\Delta u - u = D_1 g_1 + \ldots + D_n g_n$$

in the sense of distributions exists which also vanishes on $\partial\Omega$ in the sense that $u \in W^{1,2}_o(\Omega)$. Hint: minimize the functional

$$Q(u) = \int_\Omega \left\{ \frac{1}{2}|\nabla u|^2 + \frac{1}{2}u^2 + \vec{g} \cdot \nabla u \right\} dx,$$

where $\vec{g} = (g_1, \ldots, g_n)$.

Baire Category

Baire's theorem, The Uniform Boundedness Principle, The Open Mapping Theorem, The Closed Graph Theorem.

Textbooks to look at: Both of Rudin's books[4, 5] have a chapter devoted to the Baire category theorem and its applications.

51. Baire's theorem

Let (X, d) be a metric space. Recall that $V \subset X$ is called *dense* if every open $O \subset X$ intersects D; equivalently, for any point $x \in X$ and any neighborhood $U \ni x$ there is a point $p \in O \cap U$.

Theorem 49 (Baire)**.** *Let (X, d) be a complete metric space, and let $V_n \subset X$ be a sequence of dense and open subsets. Then $W = \cap_{n=1}^{\infty} V_n$ is also dense in X.*

The proof will be presented in class and can be found in [4, theorem 5.6].

Some terminology: a set $A \subset X$ is *nowhere dense* if $A \cap O$ is not dense in O for any open $O \subset X$. Equivalently, $A \subset X$ is nowhere dense if its closure has empty interior.

A set $A \subset X$ is said to be of the *first category in X* if it is the countable union of nowhere dense sets ($A = \cup_{n=1}^{\infty} A_n$, with A_n nowhere dense in X.)

A set is of the *second category* if it is not of the first category.

Baire's theorem says that the complement of a first category set $A \subset X$ is dense in X.

Typical consequences of Baire's theorem are the existence of nowhere differentiable but continuous functions (see exercise 14 of [4, chapter 5]), or the existence of continuous functions whose Fourier series do not converge (see [4, chapter 5] and also §72.2 in these notes). See also exercises 13 and 21 in [4, chapter 5]. Other examples of the kind of thing you can use Baire's theorem for are:

Exercise 94. Let $f : [0, 1] \to \mathbb{R}$ have derivatives of all orders. Suppose that for every $x \in [0, 1]$ some derivative of f vanishes, i.e. $\exists n(x) \in \mathbb{N}$ such that $f^{(n(x))}(x) = 0$. Then f is a polynomial.

Exercise 95. *Given:* a sequence of nonnegative continuous functions $f_n \in C([0,1])$ such that for each $x \in [0,1]$

$$M_x \stackrel{\text{def}}{=} \sup_{n\in\mathbb{N}} f_n(x) < \infty.$$

Prove: there exist $M < \infty$ and a nonempty interval $(a,b) \subset [0,1]$ such that $f_n(x) \leq M$ for all $n \in \mathbb{N}$ and $a < x < b$.

Exercise 96. Let X be a Banach space.

(i) Let $K \subset X$ be convex and symmetric subset. If K has nonempty interior then K contains a neighborhood of the origin.

(ii) Let $K \subset X$ be convex, closed and symmetric subset such that $X = \cup_{n=1}^{\infty} nK$. Then K contains a neighborhood of the origin. (By definition $nK = \{nx : x \in K\}$.)

52. The Uniform Boundedness Principle

The following is known as the Uniform Boundedness Principle, or as the Banach-Steinhaus theorem.

Theorem 50. *Let $\{T_a : a \in \mathcal{A}\}$ be a family of bounded linear mappings $T_a : X \to Y$, where X is a Banach space and Y is a normed vector space. If*

$$M_x \stackrel{\text{def}}{=} \sup_{a\in\mathcal{A}} \|T_a x\|_Y < \infty$$

for every $x \in X$, then

$$\sup_{a\in\mathcal{A}} \|T_a\|_{\mathcal{L}(X,Y)} < \infty.$$

Proof. Let

$$K \stackrel{\text{def}}{=} \{x \in X : \sup_{a\in\mathcal{A}} \|T_a x\|_Y \leq 1\}.$$

Our hypothesis implies that $x \in nK$ for any $n > M_x$, so that $x \in \cup_{n\in\mathbb{N}} nK$. Furthermore K is convex, closed and symmetric. By exercise 96 K contains a neighborhood of the origin, so that for some $\varepsilon > 0$ one has $\|T_a x\| \leq 1$ is $\|x\| \leq \varepsilon$. This implies $\|T_a\| \leq \varepsilon^{-1}$ where $a \in \mathcal{A}$ is arbitrary. □

Exercise 97. Let X be a Banach space and consider a sequence $\lambda_i \in X^*$. Assume $\lambda_i \stackrel{*}{\rightharpoonup} \Lambda$ as $i \nearrow \infty$. Then the sequence λ_i is bounded, i.e. $\sup_{i\in\mathbb{N}} \|\lambda_i\|_{X^*} < \infty$.

Exercise 98. Do problem 79 (ii) again, this time using the Banach-Steinhaus theorem.

53. The Open Mapping Theorem

Theorem 51 (The Open Mapping Theorem)**.** *Let $T : X \to Y$ be a bounded linear mapping between Banach spaces X and Y. If T is surjective, then $V = \{Tx : \|x\|_X \leq 1\}$ contains an open neighborhood of the origin in Y.*

For the proof see [4, theorem 5.9]. A direct consequence is

Theorem 52 (Bounded Inverse Theorem)**.** *If X and Y are Banach spaces and $T : X \to Y$ is a* bijective *bounded linear mapping, then $T^{-1} : Y \to X$ is also bounded.*

Proof. Since $V = \{Tx : \|x\|_X \leq 1\}$ contains $B_Y(0,\varepsilon) = \{y \in Y : \|y\| < \varepsilon\}$ for some $\varepsilon > 0$ we see that

$$T^{-1}\left(B_Y(0,\varepsilon)\right) \subset B_X(0,1),$$

which implies that T^{-1} is indeed bounded with $\|T^{-1}\| \leq \dfrac{1}{\varepsilon}$. □

54. The Closed Graph Theorem

The graph of a linear mapping $T : X \to Y$ is by definition

$$\mathrm{Graph}(T) = \{(x, Tx) \in X \oplus Y : x \in X\}.$$

The direct sum $X \oplus Y$ is a Banach space and the graph of T is a linear subspace.

Theorem 53 (Closed Graph Theorem)**.** *If X and Y are Banach spaces and if the graph of $T : X \to Y$ is a closed subspace of $X \oplus Y$ then T is bounded.*

Proof. Let $Z = \mathrm{Graph}(T)$. Then since Z is a closed subspace of $X \oplus Y$ it is a Banach space. Both projections $p_X : X \oplus Y \to X$, $p_Y : X \oplus Y \to Y$ given by

$$p_X(x,y) = x, \quad p_Y(x,y) = y$$

are bounded.

The projection $p_X|_Z : Z \to X$ is one-to-one and onto X, so by the bounded inverse theorem its inverse $(p_X|_Z)^{-1} : X \to Z$ is bounded.

The operator T is given by $T = p_Y \circ (p_X|_Z)^{-1}$, and hence, being the composition of bounded operators, is itself also bounded. □

Bounded Operators

Examples of Operators; Inverses and the Neumann series; A nonlinear digression; The adjoint of an operator; Kernel and Cokernel; Compact Operators

Text books to look at: The theory of the adjoint operator, and the Riesz-theory of compact operators on a Banach space is treated in much greater detail in Rudin's book on Functional Analysis [5].

55. Examples of Operators

Finite and infinite matrices

Linear maps between finite dimensional spaces are represented by matrices, i.e. a map $T : \mathbb{R}^n \to \mathbb{R}^m$ is specified by a matrix $\{T_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ via

$$(Tx)_i = \sum_{j=1}^{n} T_{ij} x_j. \tag{21}$$

The most direct generalization are linear maps on ℓ^p spaces specified by infinite matrices. Let $\{T_{ij} : i, j \in \mathbb{N}\}$ be an infinite matrix.

Exercise 99.
(i) Suppose

$$\sup_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} |T_{ij}| = M < \infty. \tag{22}$$

Show that $(Tx)_i = \sum_{j \in \mathbb{N}} T_{ij} x_j$ defines a bounded linear map $T : \ell^\infty \to \ell^\infty$
(ii) Suppose

$$\sup_{j \in \mathbb{N}} \sum_{i \in \mathbb{N}} |T_{ij}| = M' < \infty. \tag{23}$$

Show that $(Tx)_i = \sum_{j \in \mathbb{N}} T_{ij} x_j$ defines a bounded linear map $T : \ell^1 \to \ell^1$

Integral operators

The next step in generalizing matrices is to consider "continuous matrices." Let (Ω, Σ, μ) be a measure space and let $T(x, y)$ be a measurable function on $\Omega \times \Omega$.

Then one can attempt to define a linear operator by

$$Tf(x) \stackrel{\text{def}}{=} \int_\Omega T(x,y) f(y)\, d\mu(y). \tag{24}$$

Whenever this integral makes sense the resulting operator is called an integral operator with kernel $T(x,y)$. It is surprisingly hard to decide for just any kernel $T(x,y)$ if it defines a bounded operator on some L^p space. In fact no necessary *and* sufficient conditions seem to be known. The following exercises indicate some sufficient conditions or examples of bounded integral operators. See [9] for a book only about integral operators on L^2 spaces.

The first thing that comes to mind when you want to find out if $Tf \in L^p$ is to apply Hölder's inequality as often as necessary to estimate $\|Tf\|_{L^p}$ in terms of $\|f\|_{L^p}$. This works for the following class of kernels:

Exercise 100. Assume

$$\mathrm{N}_p(T) \stackrel{\text{def}}{=} \left\{ \left(\int_\Omega |T(x,y)|^{p/(p-1)}\, d\mu(y) \right)^{p-1} d\mu(x) \right\}^{1/p} < \infty.$$

Show that for any $f \in L^p(\Omega, \Sigma, \mu)$ the integral (24) exists for μ almost every $x \in \Omega$, and that (24) defines a bounded linear operator $T : L^p \to L^p$ with

$$\|Tf\|_{L^p} \le \mathrm{N}_p(T) \|f\|_{L^p}.$$

Integral operators for which

$$\mathrm{N}_2(T) = \int_{\Omega\times\Omega} |T(x,y)|^2\, dx\, dy < \infty$$

are called *Hilbert-Schmidt operators.*

The condition $\mathrm{N}_p(T) < \infty$ is sufficient, but far from necessary for a kernel T to define a bounded operator on $L^p(\Omega)$, as the following examples show.

Exercise 101. Let $f \in L^1(\mathbb{R}^n)$. Show that the operator

$$Tg(x) = f * g(x)$$

is bounded on $L^p(\mathbb{R}^n)$.

Identify the kernel $T(x,y)$ of this operator and compute $\mathrm{N}_p(T)$ as defined in exercise 100.

Exercise 102. Show that the operator

$$Tf(x) = \frac{1}{x} \int_0^x f(t)\, dt$$

is bounded from $L^p(0,1)$ to $L^p(0,1)$ for $1 < p \le \infty$.

Show that T is not bounded when $p = 1$.

Exercise 103. Show that

$$Tf(x) = \int_0^\infty \frac{f(y)\, dy}{x+y}$$

defines a bounded operator on $L^p(0,\infty)$ if $1 < p < \infty$, but not for $p = 1$ or $p = \infty$.

Exercise 104. Find a kernel $T(x,y)$ which defines a bounded integral operator on $L^{1999}(\mathbb{R})$, but not on $L^p(\mathbb{R})$ for any $p \neq 1999$. (Hint: try $T(x,y) = f(x)g(y)$.)

Finally, many ("most") bounded operators cannot be written as integral operators.

Exercise 105. Let $m : \Omega \to \mathbb{R}$ be measurable. Show that the multiplication operator

$$Mf(x) \stackrel{\text{def}}{=} m(x)f(x)$$

is bounded on $L^p(\Omega)$ if and only if $m \in L^\infty(\Omega)$.

Exercise 106. Let $Tf(x) = f(x+1)$. Then T is a bounded linear operator on $L^p(\mathbb{R})$. Show that there is no kernel $\tilde{T} : \mathbb{R}^2 \to \mathbb{R}$ with $\tilde{T} \in L^1_{\text{loc}}(\mathbb{R}^2)$ such that T is the operator with kernel $\tilde{T}$. (Hint: consider $\int_{\mathbb{R}} g(x)Tf(x)\,dx$ for characteristic functions f and g.)

56. Inverses and the Neumann series

An operator $T : X \to Y$ is invertible if it is one-to-one and onto. By the bounded inverse theorem the inverse $T^{-1} : Y \to X$ is again a bounded linear operator.

Theorem 54. *The set of bounded linear operators $T \in \mathcal{L}(X,Y)$ which are invertible is open in $\mathcal{L}(X,Y)$.*

Proof. We first use the geometric series to show that a neighborhood of the identity in $\mathcal{L}(X)$ consists of invertible operators.

If $\|T\|_{\mathcal{L}(X)} < 1$ for some $T \in \mathcal{L}(X)$ then $I - T : X \to X$ is invertible and its inverse is given by

$$(I-T)^{-1} = \sum_{k=0}^{\infty} T^k. \tag{25}$$

Indeed if $a = \|T\|_{\mathcal{L}(X)} < 1$ then the norm of the kth term in (25) is bounded by $\|T^k\| \le \|T\|^k = a^k$ so that the series in (25) is absolutely convergent and hence norm convergent in $\mathcal{L}(X)$. If one denotes the sum by $S = I + T + T^2 + \ldots$, then one has

$$TS = ST = T + T^2 + T^3 + \ldots = S - I$$

which after rearrangement gives $(I-T)S = S(I-T) = I$; so (25) is indeed true.

Next we treat the general case: Let $T : X \to Y$ be invertible, and let $S \in \mathcal{L}(X,Y)$ be some other operator for which

$$\|T - S\|_{\mathcal{L}(X,Y)} \cdot \|T^{-1}\|_{\mathcal{L}(Y,X)} = a < 1.$$

Then $T^{-1}S : X \to X$ is invertible since

$$\|T^{-1}S - I\|_{\mathcal{L}(X)} = \|T^{-1}(S-T)\|_{\mathcal{L}(X)} \le \|T-S\|_{\mathcal{L}(X,Y)} \cdot \|T^{-1}\|^{\mathcal{L}(Y,X)} < 1.$$

For a similar reason $ST^{-1} : Y \to Y$ is also invertible.

The operator $L = (T^{-1}S)^{-1}T^{-1}$ is a left inverse for S since $LS = (T^{-1}S)^{-1}T^{-1}S = I_X$; similarly, the operator $R = T^{-1}(ST^{-1})^{-1}$ is a right inverse for S.

It follows that $R = L$ and that S is invertible if

$$\|T - S\|_{\mathcal{L}(X,Y)} < \frac{1}{\|T^{-1}\|_{\mathcal{L}(Y,X)}}.$$

□

The theorem says that if an operator S is close to an invertible operator T then S must also be invertible. Here closeness is meant with the respect to the operator norm, i.e. $\|S - T\|$ should be a small number. The following example shows that this condition is stronger than one might think (i.e. operators which seem close are actually not close to each other in the operator norm.)

Exercise 107. Let $T_\varepsilon : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ be the operator given by $T_\varepsilon f = \varphi_\varepsilon * f$, in which $\varphi_\varepsilon(x) = \varepsilon^{-n}\varphi(x/\varepsilon)$ is again our favorite family of smooth compactly supported functions.

We have seen that $\|T_\varepsilon f - f\|_{L^p(\mathbb{R}^n)} \to 0$ as $\varepsilon \searrow 0$ for any $f \in L^p(\mathbb{R}^n)$, so in some sense one can say that the T_ε converge to the identity on $L^p(\mathbb{R}^n)$.

Is it true that $\|T_\varepsilon - I\|_{\mathcal{L}(L^p(\mathbb{R}^n))} \to 0$ as $\varepsilon \searrow 0$?

Hint: For the case $p = 1$ consider $T_\varepsilon(\varphi_\delta) - \phi_\delta$ and let $\delta \to 0$. For $p \neq 1$ replace ϕ_δ by $c_\delta\varphi_\delta$ where you choose $c_\delta > 0$ so that $\|c_\delta\varphi_\delta\|_{L^p} = 1$.

Exercise 108. Show that for small enough $\mu \in \mathbb{R}$ the integral equation

$$f(x) + \mu \int_0^1 \sin(x^2 - \xi^2) f(\xi)\, d\xi = g(x)$$

has a unique solution $f \in L^p(0,1)$ for any $g \in L^p(0,1)$.

Give a numerical estimate for how small μ must be.

57. A nonlinear digression: The Contraction Mapping Principle

The theorem on inverting $I - T$ for $\|T\| < 1$ is actually a special case of a nonlinear theorem.

Theorem 55. *Let (X, d) be a complete metric space, and let $F : X \to X$ be a* contraction, *i.e. for some $\theta < 1$ one has*

$$\forall_{x,y\in X} \quad d(F(x), F(y)) \leq \theta d(x, y). \tag{26}$$

Then F has a unique fixed point, i.e. there is a unique $x \in X$ such that $F(x) = x$.

Proof. Choose any $x_0 \in X$ and define inductively $x_{n+1} = F(x_n)$. Then

$$d(x_{n+1}, x_n) \leq \theta d(x_n, x_{n-1}),$$

so, by induction

$$d(x_{n+1}, x_n) \leq \theta^n d(x_1, x_0).$$

Hence, for $n < m < \infty$ one has

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n+1}) + d(x_{n+1}, x_{n+2}) + \ldots + d(x_{m-1}, x_m) \\ &\leq d(x_1, x_0)\left(\theta^n + \ldots + \theta^{m-1}\right) \\ &\leq d(x_1, x_0)\frac{\theta^n}{1-\theta}. \end{aligned}$$

It follows that x_n is a Cauchy sequence. Let x_* be its limit. Then

$$F(x_*) = F\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = x_*,$$

so a fixed point does exist.

The fixed point must be unique for if x and y are both fixed points then one has

$$d(x, y) = d(F(x), F(y)) \leq \theta d(x, y).$$

In view of $\theta < 1$ this can only hold if $d(x, y) = 0$, i.e. if $x = y$. ∎

Theorem 56. *Let X be a Banach space, and let $F : X \to X$ be a contraction (as in (26)). Then the map $\Phi : X \to X$ given by $\Phi(x) = x - F(x)$ is a homeomorphism, and its inverse $\Phi^{-1} : X \to X$ is Lipschitz continuous with constant $(1-\theta)^{-1}$, i.e.*

$$\|\Phi^{-1}(x) - \Phi^{-1}(y)\| \leq \frac{1}{1-\theta}\|x - y\|.$$

Proof. To solve $\Phi(x) = u$ for any $u \in X$ one must find a solution of $x = F(x) + u$. In other words, one must find a fixed point of the map $G(x) = F(x) + u$. From

$$\|G(x) - G(y)\| = \|F(x) - F(y)\| \leq \theta\|x - y\|$$

one sees that G is a contraction, so a solution to $x = F(x)+u$ exists: Φ is surjective. The solution to $x = F(x) + u$ is also unique, so Φ is injective.

Thus $\Phi^{-1} : X \to X$ is well defined. We now estimate $\Phi^{-1}(u) - \Phi^{-1}(v)$ for $u, v \in X$. Let

$$x = \Phi^{-1}(u), \quad y = \Phi^{-1}(v),$$

so that

$$x = F(x) + u, \quad y = F(y) + v.$$

One then has

$$\|x - y\| = \|F(x) + u - F(y) - v\| \leq \|F(x) - F(y)\| + \|u - v\| \leq \theta\|x - y\| + \|u - v\|.$$

One solves this for $\|x - y\|$ with result

$$\|x - y\| \leq \frac{\|u - v\|}{1-\theta},$$

as claimed. □

A standard application to Ordinary Differential Equations

We consider the initial value problem for a system of differential equations

$$\frac{dx}{dt} = f(x), \quad x(0) = x_0. \tag{27}$$

Here $f : \mathbb{R}^n \to \mathbb{R}^n$ is some (nonlinear) mapping, and we look for a solution $x : [0, T] \to \mathbb{R}^n$, for some $T > 0$.

Theorem 57. *Assume that f is Lipschitz continuous with*

$$\forall_{x,y \in \mathbb{R}^n} |f(x) - f(y)| \leq L|x - y|.$$

Let $0 < T < L$ be given. Then (27) has a unique solution $x \in C([0, T]; \mathbb{R}^n)$ for any given initial data $x_0 \in \mathbb{R}^n$.

Proof. Rewrite (27) as an integral equation,

$$x(t) = x_0 + \int_0^t f(x(s))\, ds. \tag{28}$$

Denote the right hand side of this equation by $F(x)$, more precisely, for any $x \in C([0, T]; \mathbb{R}^n)$ we put

$$(Fx)(t) \stackrel{\text{def}}{=} x_0 + \int_0^t f(x(s))\, ds, \text{ for } t \in [0, T].$$

Since $LT < 1$ this map is a contraction on the Banach space $X = C([0,T];\mathbb{R}^n)$:

$$\begin{aligned}\sup_{0\le t\le T}|(Fx)(t)-(Fy)(t)| &= \sup_{0\le t\le T}\left|\int_0^t \{f(x(s))-f(y(s))\}\,ds\right| \\ &\le \sup_{0\le t\le T}\int_0^t |f(x(s))-f(y(s))|\,ds \\ &\le T\sup_{0\le t\le T}|f(x(t))-f(y(t))| \\ &\le LT\sup_{0\le t\le T}|x(t)-y(t)|.\end{aligned}$$

The unique fixed point $x \in C([0,T];\mathbb{R}^n)$ of F is the solution of (28).

At this point we only know that x is a continuous function of t (since we constructed x as an element of $C([0,T];\mathbb{R}^n)$), but (28) implies that x is actually continuously differentiable, and the fundamental theorem of calculus implies that x actually satisfies (27). □

58. The adjoint of a bounded operator

Let $T : X \to Y$ be a bounded linear operator between Banach spaces. If $\mu \in Y^*$ is a bounded functional on Y then one defines a bounded linear functional $T^*\mu$ on X by

$$T^*\mu(x) \stackrel{\text{def}}{=} \mu(Tx).$$

In this way a linear transformation $T^* : Y^* \to X^*$ is defined. This transformation is called the *adjoint of the operator* T. Clearly one has

$$\begin{aligned}\|T^*\mu\| &= \sup_{\|x\|\le 1}|T^*\mu(x)| \\ &= \sup_{\|x\|\le 1}|\mu(Tx)| \\ &\le \|\mu\|\sup_{\|x\|\le 1}\|Tx\| \\ &= \|T\|\,\|\mu\|\end{aligned}$$

so that T^* is bounded and $\|T^*\|_{\mathcal{L}(Y^*,X^*)} \le \|T\|_{\mathcal{L}(X,Y)}$.

Exercise 109. Use the Hahn-Banach theorem to show that

$$\|T^*\|_{\mathcal{L}(Y^*,X^*)} = \sup_{\substack{\|x\|\le 1\\ \|\mu\|\le 1}}\mu(Tx) = \|T\|_{\mathcal{L}(X,Y)}.$$

Exercise 110.

(i) Let $T : X \to Y$ be a bounded linear operator, and let $\{x_i \in X : i \in \mathbb{N}\}$ be a weakly convergent sequence, $x_k \rightharpoonup x_\infty$. Show that $T(x_k)$ is weakly convergent in Y, and that $Tx_k \rightharpoonup Tx_\infty$.

(ii) Let $\{\mu_i \in X^* : i \in \mathbb{N}\}$ be a weak* convergent sequence, $\mu_k \stackrel{*}{\rightharpoonup} \mu_\infty$. Show that $T^*(\mu_k)$ is weak* convergent in X^*, and that $T^*\mu_k \stackrel{*}{\rightharpoonup} T^*\mu_\infty$.

Exercise 111. Let $f \in L^1(\mathbb{R}^n)$. Find the adjoint of the operator on $L^p(\mathbb{R}^n)$ given by $Tg = f * g$.

Exercise 112. Show that the operator S given by

$$Sf(x) = \int_x^1 \frac{f(\xi)}{\xi}\, d\xi$$

is bounded on $L^p(0,1)$ for all $p \in [1,\infty)$, but not for $p = \infty$.

Show that S is the adjoint of the operator T in exercise 102.

59. Kernel and Cokernel

We consider a bounded operator $T : X \to Y$ where X and Y are Banach spaces. The *kernel of* T is defined to be

$$\ker T = \{x \in X : Tx = 0\}.$$

From linear algebra we know that T is one-to-one if and only if $\ker T = \{0\}$.

The *range of* T is by definition

$$R(T) = \{Tx : x \in X\} \subset Y.$$

The range of an operator is a linear subspace of Y which in general does not have to be closed. If the range is closed however, then one defines the *cokernel of* T to be the following quotient of Banach spaces

$$\operatorname{coker} T = Y/R(T).$$

If $R(T)$ is closed, then T is surjective if and only if its cokernel is trivial, i.e. $\operatorname{coker} T = \{0\}$.

Exercise 113. Show that the following operators do not have closed range:
(i) $X = Y = L^p(0,1)$, and $Tf(x) = xf(x)$.
(ii) $X = Y = \ell^p$ and $(Tx)_j = 2^{-j}x_j$.
(iii) $X = Y = L^p(0,1)$ and $Tf(x) = \int_0^x f(\xi)\, d\xi$.

There is a simple relation between the kernels and cokernels of an operator T and its adjoint T^*.

Theorem 58. (i) *Let $T : X \to Y$ be bounded. Then*

$$R(T)^\perp = \ker T^* \text{ and } {}^\perp R(T^*) = \ker T.$$

(ii) *If $R(T)$ is closed then one has*

$$R(T) = {}^\perp \ker T.$$

Here for any linear subspaces $L \subset X$ and $M \subset X^*$ one defines the so-called *annihilators*

$$L^\perp = \{\lambda \in X^* : \forall_{x \in L} \lambda(x) = 0\}, \tag{29}$$

$${}^\perp M = \{x \in X : \forall_{\mu \in M} \mu(x) = 0\}. \tag{30}$$

I leave the proof as an exercise (but see [5, theorem 4.12] for part (i); to prove part (ii) use the Hahn Banach theorem.)

60. Compact operators

Let X, Y be Banach spaces. By definition an operator $T : X \to Y$ is compact if the image of the unit ball in X under T has compact closure in Y, i.e. $\{Tx : \|x\| \leq 1\}$ has compact closure in Y.

The following is an equivalent definition: *T is compact if every bounded sequence $x_n \in X$ has a subsequence for which $Tx_n \in Y$ converges in Y.*

The set of $T \in \mathcal{L}(X, Y)$ which are compact is denoted by $\mathcal{K}(X, Y)$.

Exercise 114.

(i) Show that if $T : X \to Y$ is invertible, with $T^{-1} : Y \to X$ also bounded, then T compact implies X finite dimensional.

(ii) Show that if $T, S : X \to Y$ are compact operators, then for any $\lambda, \mu \in \mathbb{R}$ the operator $\lambda T + \mu S$ is also a compact operator; in other words show that $\mathcal{K}(X, Y)$ is a linear subspace of $\mathcal{L}(X, Y)$.

Theorem 59. *If T_n is a sequence of compact operators from X to Y, and if T_n converges in the operator norm to T, then T is also compact.*

In other words, compact operators form a closed subset of the set of bounded operators.

Proof. Let $x_i \in X$ be a bounded sequence. We must show that Tx_i has a convergent subsequence. Each T_n is compact, so for each fixed n one can find a subsequence x_{i_j} such that $\lim_{j\to\infty} T_n x_{i_j} \in Y$ exists. Using Cantor's diagonalization trick one can find one subsequence such that

$$\lim_{j\to\infty} T_n x_{i_j} = y_n \in Y$$

exists for all $n \in \mathbb{N}$.

We now show that $\{Tx_{i_j} : j \in \mathbb{N}\}$ is a Cauchy sequence in Y, and hence a convergent subsequence of the Tx_i. Let $\varepsilon > 0$ be given. Choose some $n_0 \in \mathbb{N}$ such that $\|T_{n_0} - T\| < \varepsilon/3$. Since $\{T_{n_0} x_{i_j} : j \in \mathbb{N}\}$ converges in Y as $j \nearrow \infty$ it is also a Cauchy sequence. Hence an $N(\varepsilon) \in \mathbb{N}$ exists for which $\|T_{n_0} x_{i_j} - T_{n_0} x_{i_k}\|_Y < \varepsilon/3$ for all $j, k \geq N(\varepsilon)$. One then also has

$$\begin{aligned} \|Tx_{i_j} - Tx_{i_k}\|_Y &\leq \|Tx_{i_j} - T_{n_0} x_{i_j}\|_Y + \|T_{n_0} x_{i_j} - T_{n_0} x_{i_k}\|_Y + \|T_{n_0} x_{i_k} - Tx_{i_k}\|_Y \\ &< \varepsilon/3 + \varepsilon/3 + \varepsilon/3 \\ &= \varepsilon. \end{aligned}$$

for all $j, k \geq N(\varepsilon)$. □

Exercise 115. If $T : X \to Y$ and $S : Y \to Z$ are bounded operators, and if either T or S is compact, then show that $ST : X \to Z$ is also compact.

Theorem 60. *Let $\lambda \in \mathbb{C} \setminus \{0\}$. If $T : X \to X$ is compact, then*

(i): $\ker(\lambda I - T)$ *is finite dimensional;*

(ii): $R(\lambda I - T)$ *is closed;*

(iii): $R(\lambda I - T)$ *has finite codimension, and* $\dim \ker(\lambda I - T) = \operatorname{codim} R(\lambda I - T)$.

Proof. We only prove (i) here. For the rpoof of (ii) and (iii) you should look at Rudin's *Functional Analysis* [5].

Let $x_n \in \ker(\lambda I - T)$ be a bounded sequence. Then for some subsequence Tx_{n_i} converges. Since $Tx_n = \lambda x_n$ we conclude that $x_{n_i} = \lambda^{-1} T x_{n_i}$ also converges.

Thus every bounded sequence in $\ker(\lambda I - T)$ has a convergent subsequence. This implies that the closed unit ball in $\ker(\lambda I - T)$ is compact, and hence the Banach space $\ker(\lambda I - T)$ is finite dimensional, by Lemma 7.

□

61. Finite rank operators.

An operator $T : X \to Y$ has finite rank if its range is $R(T)$ finite dimensional. For any vectors $y_1, \ldots, y_m \in Y$ and functionals $\lambda_1, \ldots, \lambda_m \in X^*$ the linear operator given by

$$Tx = \sum_{i=1}^{m} \lambda_i(x) y_i \tag{31}$$

has finite rank since its range is contained in the subspace of Y spanned by the y_i.

Lemma 61. *Any finite rank operator is of the form* (31).

Proof. Let $y_1, \ldots, y_m \in Y$ be a basis for $R(T)$. Then any vector $y \in R(T)$ is of the form $y = c_1 y_1 + \ldots + c_m y_m$ where the c_i depend continuously on $y \in R(T)$. The function which assigns $c_i \in \mathbb{R}$ to $y \in R(T)$ thus defines a bounded linear functional $c_i : R(T) \to \mathbb{R}$, and one has

$$y = c_1(y) y_1 + \ldots + c_m(y) y_m$$

for all $y \in R(T)$. In particular one has

$$Tx = c_1(Tx) y_1 + \ldots + c_m(Tx) y_m$$

for all $x \in X$. Hence T has the form (31) if one defines $\lambda_i(x) = c_i(Tx)$. □

Theorem 62. *A finite rank operator is compact.*

Proof. For any bounded sequence $x_n \in X$ the sequence Tx_n is a bounded sequence in $R(T)$. Since $R(T)$ is finite dimensional any bounded sequence in $R(T)$ has a convergent subsequence. □

Exercise 116.
(i) Consider the operator $T : L^p(0,1) \to L^p(0,1)$ given by

$$Tf(x) = \int_0^1 \sin(x^2 - \xi^2) f(\xi)\, d\xi.$$

Show that T has finite rank, and write T in the form (31). (Hint: $\sin(\alpha + \beta) = ?$)
(ii) For which $\mu \in \mathbb{C}$ does the integral equation

$$f(x) + \mu \int_0^1 \sin(x^2 - \xi^2) f(\xi)\, d\xi = g(x)$$

have a unique solution $f \in L^2(0,1)$ for all $g \in L^2(0,1)$.

62. Compact integral operators.

Let $T : \Omega \times \Omega \to \mathbb{R}$ be a measurable function whose norm

$$\mathrm{N}_p(T) = \left\{ \left(\int_\Omega |T(x,y)|^{p/(p-1)}\, d\mu(y) \right)^{p-1} d\mu(x) \right\}^{1/p}$$

is finite. Such a function defines a bounded integral operator on $L^p(\Omega)$ (see exercise 100).

Theorem 63. *If $1 < p < \infty$ then an integral operator with kernel $T(x,y)$ for which $\mathrm{N}_p(T) < \infty$ is compact.*

Proof. We only do the case $p = 2$. This case is easier since the $\mathrm{N}_p(T)$ norm is then given by the more familiar quantity

$$\mathrm{N}_p(T)^2 = \int_{\Omega\times\Omega} |T(x,y)|^2 \, dx \, dy.$$

In measure theory it is shown that "simple" functions of the form

$$\tilde{T}(x,y) = \sum_{i=1}^{m} c_i \chi_{E_i}(x) \chi_{F_i}(y)$$

are dense in $L^2(\Omega \times \Omega)$. Thus the integral operator on $L^2(\Omega)$ with kernel T can be approximated by integral operators whose kernels are of the form $\tilde{T}$. But an operator with kernel $\tilde{T}$ has finite rank and is therefore compact. So the operator with kernel T is a limit of compact operators and therefore is itself also compact.

To prove the general case one must show that the quantity $\mathrm{N}_p(T)$ is a norm and that any kernel T with $\mathrm{N}_p(T) < \infty$ can be approximated by functions of the form $\tilde{T}$. □

Not all integral operators are compact.

Exercise 117. Let $T : L^p(\mathbb{R}^n) \to L^p(\mathbb{R}^n)$ be given by convolution with an $f \in L^1(\mathbb{R}^n)$, i.e. $Tg = f * g$. Show that T is not compact (unless $T = 0$; if you wish you can simplify the problem by assuming that f is continuous with compact support.)

Another example of a bounded integral operator which is not compact is the operator

$$\text{(32)} \qquad \begin{cases} Tf(x) = \displaystyle\int_0^1 \tilde{T}(x,y) f(y) \, dy, \\ \tilde{T}(x,y) = \sin 2n\pi x, \text{ for } 2^{-n-1} \le y < 2^{-n}. \end{cases}$$

This operator is bounded from $L^1(\Omega)$ to $L^1(\Omega)$, with $\Omega = (0,1)$ the unit interval, but since $T\varphi_n = \sin 2n\pi x$, for $\varphi_n(x) = 2^{n+1}\chi_{(2^{-n-1},2^{-n})}(x)$, T cannot be compact.

Exercise 118. Show that the operator T defined above in (32) is bounded *and compact* from $L^p(0,1)$ to $L^p(0,1)$ for $1 < p < \infty$.

We haven't changed the kernel $\tilde{T}$ so we still have $T\varphi_n = \sin 2n\pi x$ for all $n = 1, 2, \ldots$. Hence $T\varphi_n$ does not have a convergent subsequence in $L^p(0,1)$. Why does this not show that $T : L^p \to L^p$ is not compact?

63. Green's operator

In section 50.4 we saw that for any *bounded* domain $\Omega \subset \mathbb{R}^n$ and any $f \in L^2(\Omega)$ there is a unique solution $u \in W_o^{1,2}(\Omega)$ of the equation

$$\text{(33)} \qquad -\Delta u = f,$$

in the sense of distributions. The solution u was obtained by minimizing

$$Q(u) = \int_\Omega \{\tfrac{1}{2}|\nabla u|^2 - f(x)u(x)\} \, dx$$

over all $u \in W_o^{1,2}(\Omega)$.

Let us denote the map which sends $f \in L^2(\Omega)$ to $u \in W_o^{1,2}(\Omega)$ by $G : L^2(\Omega) \to W_o^{1,2}(\Omega)$.

Theorem 64. *The map G is bounded, linear and compact.*

Proof. Linearity follows from the fact that equation (33) is linear. If $Gf = u$ and $Gg = v$ then $f = -\Delta u$, and $g = -\Delta v$, so that $w = \alpha u + \beta v \in W_o^{1,2}(\Omega)$ satisfies $-\Delta w = \alpha f + \beta g$. Hence $w = G(\alpha f + \beta g)$.

To prove boundedness we argue as follows. Since u is the minimizer of $Q(u)$ we have

$$\left.\frac{dQ(tu)}{dt}\right|_{t=1} = 0. \tag{34}$$

From

$$Q(tu) = \frac{t^2}{2}\int_\Omega |\nabla u|^2 \, dx - t\int_\Omega f(x)u(x) \, dx$$

and thus

$$\frac{dQ(tu)}{dt} = t\int_\Omega |\nabla u|^2 \, dx - \int_\Omega f(x)u(x) \, dx$$

one then concludes from (34) that

$$\int_\Omega |\nabla u|^2 \, dx = \int f(x)u(x) \, dx \leq \|u\|_{L^2}\|f\|_{L^2}. \tag{35}$$

By Poincaré's inequality (Lemma 45) we have

$$\|u\|_{L^2} \leq 2|\Omega|^{1/n}\left(\int |\nabla u|^2 dx\right)^{1/2}$$

Hence

$$\|u\|_{L^2}^2 \leq 4|\Omega|^{2/n}\int |\nabla u|^2 dx.$$

Apply this to (35) to get, after cancellation,

$$\|u\|_{L^2} \leq 4|\Omega|^{2/n}\|f\|_{L^2}.$$

Thus Green's operator is bounded from $L^2(\Omega)$ to $L^2(\Omega)$.

To prove compactness we consider any bounded sequence $f_n \in L^2(\Omega)$. Since $G : L^2 \to L^2$ is bounded Gf_n is a bounded sequence in $L^2(\Omega)$. By (35) Gf_n is also bounded in $W_o^{1,2}(\Omega)$. The Rellich-Kondrachov theorem implies that the sequence Gf_n must have a convergent subsequence in $L^2(\Omega)$. Hence G is indeed compact. □

Hilbert Spaces

Definition; Examples; The Riesz representation theorem; Orthonormal sets and bases; Examples of Orthonormal sets in $L^2(0, 2\pi)$; More examples of orthogonal sets, or "Orthogonal Polynomials 101"; The Spectral Theorem for Symmetric Compact Operators; Eigenfunctions of the Laplacian

Text books to look at: Both Rudins [4, 5] have chapters on Hilbert spaces. The spectral theorem is in Zimmer's book [8].

64. **Definition.**

A *pre-Hilbert* space is a real or complex vector space H with a positive definite inner product (x, y). In the real case this means that (x, y) is linear both in x and in y; that (x, y) is symmetric, $(x, y) = (y, x)$; and that $(x, x) > 0$ for all $x \neq 0$. If the vector space is complex one requires

$$(\lambda x, y) = \lambda(x, y)$$
$$(x, \lambda y) = \bar{\lambda}(x, y)$$
$$(x, y) = \overline{(y, x)}.$$

The quantity

$$\|x\| = \sqrt{(x, x)}$$

defines a norm on H. If H is complete with this norm, then H is called a *Hilbert Space.*

64.1. Proof that $\|x\|$ satisfies the triangle inequality

First one proves the Cauchy-Schwarz inequality,

$$|(x, y)| \leq \|x\| \, \|y\|$$

by observing that

$$P(t) = (x + ty, x + ty) = \|x\|^2 + 2\Re(x, y) + t^2\|y\|^2$$

is a nonnegative quadratic poynomial. From Math 112 we therefore know that "$b^2 - 4ac < 0$", i.e.

$$\big(2\Re(x, y)\big)^2 \leq 4\|x\|^2\|y\|^2 \text{ i.e. } \Re(x, y) \leq \|x\| \, \|y\|.$$

Replace x by $e^{i\theta}y$ to get

$$\Re\big(e^{i\theta}(x,y)\big) \le \|x\|\,\|y\|.$$

Choose θ so that $e^{i\theta}(x,y) = |(x,y)|$.

Given the Cauchy-Schwarz inequality one has

$$\begin{aligned}\|x+y\| &= \sqrt{(x+y,x+y)}\\ &= \sqrt{\|x\|^2 + 2\Re(x,y) + \|y\|^2}\\ &\le \sqrt{\|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2}\\ &= \|x\| + \|y\|.\end{aligned}$$

65. Examples of Hilbert Spaces

If (Ω,Σ,μ) is a σ finite measure space then $L^2(\Omega,\Sigma,\mu)$ with inner product

$$(f,g) = \int_\Omega f(x)\overline{g(x)}\,d\mu(x)$$

is a Hilbert space. (The complex conjugate is included, and in this section we will use the same notation for both the real and complex versions of L^2!)

If $w(x)$ is a bounded measurable function on Ω with $w(x) \ge \delta > 0$ for μ-a.e. $x \in \Omega$, then the expression

$$(f,g)_w = \int_\Omega f(x)\overline{g(x)}w(x)\,d\mu(x)$$

defines another inner product on $L^2(\Omega,\Sigma,\mu)$. Both inner products yield equivalent norms.

If $\Omega \subset \mathbb{R}^n$ is open then $W^{1,2}(\Omega)$ is also a Hilbert Space, provided one gives it the inner product

$$(f,g)_{W^{1,2}} \stackrel{\text{def}}{=} \int_\Omega \{f(x)\overline{g(x)} + \nabla f(x)\cdot\overline{\nabla g(x)}\}\,dx. \tag{36}$$

One can also modify this inner product by including a weight function w, i.e.

$$(f,g)_w \stackrel{\text{def}}{=} \int_\Omega \{f(x)\overline{g(x)} + \nabla f(x)\cdot\overline{\nabla g(x)}\}\,w(x)dx.$$

If w is measurable and if $c \le w(x) \le C$ for constants $0 < c < C$ then this expression defines an inner product on $W^{1,2}$, and the resulting norm is equivalent to the usual $W^{1,2}$ norm.

Exercise 119. Let $\Omega \subset \mathbb{R}^n$ be open with finite volume.

The quantity

$$(f,g)_o \stackrel{\text{def}}{=} \int_\Omega \nabla f(x)\cdot\overline{\nabla g(x)}\,dx.$$

does NOT define an inner product on $W^{1,2}(\Omega)$, since one has $(f,f) = 0$ for all constant functions.

Show that $(f,g)_o$ does define an inner product on $W^{1,2}_o(\Omega)$, and that the resulting norm is equivalent to the usual norm on $W^{1,2}$. *Hint:* use Poincaré's inequality (Lemma 45).

66. The Riesz representation theorem.

Lemma 65. *Let K be a nonempty closed and convex subset of a Hilbert space H. Then K contains a unique element x with minimal norm.*

More generally, given any point $p \in H$ there is a unique nearest point to p in K.

Proof. The general case follows from the special by translating p to the origin.

Let $d = \inf\{\|x\|^2 : x \in K\}$. Choose a sequence $x_n \in K$ with $\|x_n\|^2 < d + \frac{1}{n}$. Then $\{x_n : n \in \mathbb{N}\}$ is a Cauchy sequence; the limit of this Cauchy sequence must lie in K and minimizes $\|x\|$.

To see that x_n is a Cauchy sequence let $n < m \in \mathbb{N}$ be given. Then $y = (x_n + x_m)/2$ belongs to K since K is convex. One has

$$\begin{aligned} d \le \|y\|^2 &= \frac{1}{4}\{\|x_n\|^2 + 2\,(x_n, x_m) + \|x_m\|^2\} \\ &= \frac{1}{4}\{2\|x_n\|^2 + 2\|x_m\|^2 - \|x_n - x_m\|^2\} \\ &\le d + \frac{1}{n} - \frac{1}{4}\|x_n - x_m\|^2 \end{aligned}$$

and hence

$$\|x_n - x_m\|^2 \le \frac{4}{n}.$$

To see that there is only one nearest point one supposes that x and y are both nearest points and observes that $x, y, x, y, x, y, x, \ldots$ is a distance minimizing sequence, hence a Cauchy sequence by the previous arguments. Thus $x = y$. □

Corollary 66. *If $L \subset H$ is a closed linear subspace then any $x \in H$ can be written as $x = y + z$ with $y \in L$ and $z \perp L$. The components y and z are unique.*

Proof. L is closed and convex so we can apply the previous Lemma. The projection $y \in L$ is the nearest point to x contained in L. □

Corollary 67. *A linear subspace $L \subset H$ is dense if and only if $x \perp L$ implies $x = 0$.*

Proof. Suppose $x \perp L$ implies $x = 0$. Then any $x \in H$ can be written as $x = y + z$ where y is in the closure of L and z is perpendicular to the closure of L. By assumption $z = 0$ so $x = y$ belongs to the closure of L: L is dense.

Suppose L is dense in H and let $x \in H$ satisfy $x \perp L$. Choose $x_n \in H$ with $\|x_n - x\| \to 0$. Then

$$\|x\|^2 = (x, x) = \lim_{n\to\infty} (x, x_n) = 0.$$

We see that $x \perp L$ implies $x = 0$. □

The following is called the Riesz representation theorem.

Theorem 68. *For every bounded linear functional λ on H a unique $x_\lambda \in H$ exists such that*

$$\lambda(x) = (x_\lambda, x) \text{ for all } x \in H.$$

One has

$$\|\lambda\| = \|x_\lambda\|.$$

Proof. If $\lambda = 0$ then one chooses $x_\lambda = 0$. We may assume that $\lambda \neq 0$.

Let $L = \ker(\lambda)$. Since $\lambda \neq 0$ we have $H \neq L$, so a vector $v \in H \setminus L$ exists. Since L is a closed linear subspace we may replace v by its component perpendicular to L, and after normalizing we may assume that $\|v\| = 1$.

Since $v \notin L$ we have $\lambda(v) \neq 0$. Thus we can write any $x \in H$ as

$$x = l + m, \quad \ell \stackrel{\text{def}}{=} x - \frac{\lambda(x)}{\lambda(v)} v, \quad m \stackrel{\text{def}}{=} \frac{\lambda(x)}{\lambda(v)} v.$$

One sees that $\lambda(\ell) = 0$, so that $\ell \in L$. Taking the inner product with v we see that

$$(x, v) = \frac{\lambda(x)}{\lambda(v)}(v, v) = \frac{\lambda(x)}{\lambda(v)}.$$

hence

$$\lambda(x) = (x, \lambda(v)v)$$

for all $x \in H$ so that we may choose $x_\lambda = \lambda(v)v$. □

Corollary 69. *Every Hilbert space is reflexive.*

Proof. Let $\Lambda \in H^{**}$ be given, i.e. $\Lambda : H^* \to \mathbb{R}$ is a bounded linear functional.

By the Riesz representation theorem every $\varphi \in H^*$ is of the form $\varphi = \varphi_x$ for some $x \in H$, where $\varphi_x(y) \stackrel{\text{def}}{=} (x, y)$. Moreover the map $x \mapsto \varphi_x$ is an isometry of H with H^*.

The functional $\Lambda : H^* \to \mathbb{R}$ then gives us a functional $\tilde{\Lambda} : H \to \mathbb{R}$ via $\tilde{\Lambda}(x) \stackrel{\text{def}}{=} \Lambda(\varphi_x)$. The Riesz representation theorem again says that the functional $\tilde{\Lambda}$ must be of the form $\tilde{\Lambda} = \varphi_{x_0}$ for some $x_0 \in H$. We therefore have that for all $\varphi = \varphi_x \in H^*$

$$\Lambda(\varphi_x) = \tilde{\Lambda}(x) = \varphi_{x_0}(x) = (x_0, x) = (x, x_0) = \varphi_x(x_0).$$

In other words,

$$\Lambda(\varphi_x) = \widehat{x_0}(\varphi_x).$$

We conclude that every bounded linear functional on H^* is of the form $\widehat{x_0}$ for some $x_0 \in H$, so that H is reflexive. □

67. Orthonormal sets and bases

A sequence of vectors $\{x_n \in H \mid n \in \mathbb{N}\}$ is *orthogonal* if $(x_i, x_j) = 0$ for all $i \neq j$.

An orthogonal sequence is called orthonormal if it consists of unit vectors.

Lemma 70. *If $\{x_n \in H \mid n \in \mathbb{N}\}$ is an orthonormal sequence then the series $\sum_{i=1}^{\infty} a_i x_i$ converges iff $\sum |a_i|^2 < \infty$, and one has*

$$\left\| \sum_{i=1}^{\infty} a_i x_i \right\| = \sqrt{\sum |a_i|^2}.$$

Proof. Denote the partial sums of $\sum a_i x_i$ by

$$s_N \stackrel{\text{def}}{=} \sum_{i=1}^{N} a_i x_i.$$

Using the definition of the norm and the orthogonality of the x_i one computes

$$\|s_n - s_m\|^2 = \sum_{i=n+1}^{m} |a_i|^2. \tag{37}$$

If we suppose that $\sum |a_i|^2 < \infty$ then (37) implies that $\|s_n - s_m\|^2 \to 0$ as $n \to \infty$. Thus s_n is a Cauchy sequence and the partial sums s_n converge.

Conversely, suppose the sum converges. Then the sequence of norms $\|s_n\|$ also converges, and

$$\sum_{i=1}^{\infty} |a_i|^2 = \lim_{N\to\infty} \sum_{i=1}^{N} |a_i|^2 = \lim_{N\to\infty} \|s_N\|^2 < \infty.$$

□

An orthonormal sequence $\{x_i : i \in \mathbb{N}\}$ is called an orthonormal basis or a *complete orthonormal set* for H if any $x \in H$ can be written as

$$x = \sum_{i=1}^{\infty} a_i x_i.$$

Lemma 71. *An orthonormal set $\{x_i\}_{i\in\mathbb{N}}$ is complete if $x \perp x_i$ for all $i \in \mathbb{N}$ implies $x = 0$.*

Proof. Let $x \in H$ be given. The orthogonal projection onto the finite dimensional space spanned by $x_1, \ldots, x_N$ is given by

$$P_N x = a_1 x_1 + \ldots + a_N x_N,$$

where

$$a_i = (x_i, x)\,.$$

Since $\|P_N x\| \le \|x\|$ we get

$$\sum_{i=1}^{N} |a_i|^2 = \|P_N x\|^2 \le \|x\|^2.$$

This holds for all N and hence

$$\sum_{i=1}^{\infty} |a_i|^2 \le \|x\|^2.$$

(This is called *Bessel's inequality.*)

It follows that the sum $\sum_{i=1}^{\infty} a_i x_i$ converges in H □

If a Hilbert space has a complete orthonormal set $\{x_i : i \in \mathbb{N}\}$ then we can define a map $\varphi : H \to \ell^2(\mathbb{N})$ by

$$\varphi(x) = \{a_i : i \in \mathbb{N}\} \Leftrightarrow a_i = (x_i, x)$$

This map is an isomorphism of Hilbert spaces, i.e. it's a linear map, it's bijective, and it preserves the inner product.

Theorem 72. *Every separable Hilbert space has a complete orthonormal set and hence is isomorphic with $\ell^2(\mathbb{N})$.*

Proof. Let $y_i \in H$ be a dense sequence. From this sequence we extract a subsequence by discarding every y_k which is linearly dependent on $\{y_1, \ldots, y_{k-1}\}$. Denote the resulting subsequence by $\{z_i : i \in \mathbb{N}\}$. This subsequence is linearly independent. The linear subspace it spans contains all the y_i and hence is dense in H.

Now apply the Gramm-Schmidt procedure to the sequence $\{z_i\}$. In other words, let $\hat{x}_k$ be the component of z_k which is perpendicular to $z_1, \ldots, z_{k-1}$, and let $x_k = \hat{x}_k / \|\hat{x}_k\|$. Then $\{x_k : k \in \mathbb{N}\}$ is an orthonormal set. It spans the same subspace as the z_k and therefore $\{x_k : k \in \mathbb{N}\}$ is a complete orthonormal system. □

Exercise 120. If x_i, $i \in \mathbb{N}$ is an orthonormal basis for H then you can write any $x \in H$ as a convergent sum $x = \sum_i a_i x_i$. Is this sum always absolutely convergent?

68. Examples of Orthonormal sets in $L^2(0, 2\pi)$.

As you can check for yourself, the functions

$$e_n(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} e^{inx}, \text{ with } n \in \mathbb{Z}$$

form an orthonormal set in $L^2(0, 2\pi)$.

Since $e^{inx} = \left(e^{ix}\right)^n$, finite linear combinations in the e_n, i.e. expressions of the form

$$P(x) = c_{-N} e_{-N}(x) + \ldots + c_N e_N(x), \text{ with } c_j \in \mathbb{C}$$

are in fact polynomials in e^{ix} and e^{-ix}. They are called *trigonometric polynomials.*

Theorem 73 (Fourier series)**.** *$\{e_n : n \in \mathbb{Z}\}$ is a complete orthonormal system in $L^2(0, 2\pi)$.*

Proof. We have to show that the space of finite linear combinations of the e_n is dense in L^2. There are many ways to do this. One way is to observe that the Stone-Weierstrass theorem implies that any continuous 2π periodic function can be uniformly approximated by a trigonometric polynomial. Since continuous functions are dense in $L^2(0, 2\pi)$ we conclude that trigonometric polynomials are dense in L^2.

Another proof which avoids the Stone-Weierstrass theorem goes like this: one must show that if $f \in L^2(0, 2\pi)$ satisfies $(f, e_n) = 0$ for all $n \in \mathbb{Z}$, then $f = 0$. The hypothesis implies that $(f, P) = 0$ for all trigonometric polynomials $P(x)$. Now construct for any given interval $(a, b) \subset (0, 2\pi)$ a sequence of trigonometric polynomials $P_N(x)$ such that

$$\sup_{N,x} |P_N(x)| < \infty$$

$$\lim_{N \to \infty} P_N(x) = \chi_{(a,b)}(x) \text{ pointwise.}$$

The dominated convergence theorem then implies

$$\int_{(a,b)} f(x)\, dx = \lim_{N \to \infty} \int_0^{2\pi} P_N(x) f(x)\, dx = 0$$

from which one concludes that $f(x) = 0$ a.e. □

Exercise 121. Show that in the above proof one can take

$$P_N(x) = \int_0^x \{Q(t-a) - Q(t-b)\}\, dt$$

where

$$Q(t) = \frac{(1+\cos x)^N}{\int_0^{2\pi}(1+\cos x)^N\, dx}.$$

This theorem implies that for any $f \in L^2(0, 2\pi)$ one has

$$f(x) = \sum_{n\in\mathbb{Z}} \hat{f}_n e^{inx} \text{ in } L^2,$$

where

$$\hat{f}_n \stackrel{\text{def}}{=} \frac{1}{2\pi}\int_0^{2\pi} e^{-inx} f(x)\, dx.$$

The qualification "in L^2" is important: we have shown that the partial sums

$$s_N f(x) \stackrel{\text{def}}{=} \sum_{n=-N}^{N} \hat{f}_n e^{inx}$$

converge in $L^2(0, 2\pi)$ to the function $f(x)$. This does not allow us to conclude that $\lim_{N\to\infty} s_N f(x) = f(x)$ for even one $x \in [0, 2\pi]$!

Exercise 122. If you've never done this before, compute the Fourier series of the function $f(x) = \chi_{(0,\pi)}(x) - \chi_{(\pi,2\pi)}(x)$.

Assuming that $\lim_{N\to\infty} s_N f(x) = f(x)$ holds pointwise, what do you get for $x = \frac{\pi}{2}$?

Exercise 123. Compute the Fourier series of the function $f(x) = x(2\pi - x) \in L^2(0, 2\pi)$.

Assuming that $\lim_{N\to\infty} s_N f(x) = f(x)$ holds pointwise, what do you get for $x = \pi$?

Exercise 124. Let $0 < a < 1$. Compute the Fourier series of the 2π periodic function which for $-\pi < x < \pi$ is given by $f(x) = \sin ax$. (This function is discontinuous at $x = \pi + 2k\pi$).

Assume again that $\lim_{N\to\infty} s_N f(x) = f(x)$ holds pointwise. What do you get if you substitute $x = \frac{\pi}{2}$?

Exercise 125.

(i) Show that the system $\{\sin(n\pi x) : n = 1, 2, 3, \dots\}$ is complete in $L^2(0, \pi)$.

(ii) Show that the system $\{\cos(n\pi x) : n = 0, 1, 2, 3, \dots\}$ is complete in $L^2(0, \pi)$.

Exercise 126. The Walsh-system $\{W_{k,n}(x) \mid n \in \mathbb{N}, 0 \le k < 2^n\}$ is defined by

$$W_{1,1}(x) = 1,$$

$$W_{k,n}(x) = \begin{cases} 1 & \text{for } k2^{-n} < x < (k+1/2)2^{-n}, \\ -1 & \text{for } (k+1/2)2^{-n} < x < (k+1)2^{-n}, \end{cases}$$

Show that this system is complete.

69. More examples of orthogonal sets, or "Orthogonal Polynomials 101"

Let $w(x) > 0$ be an integrable function on an interval $(a, b) \subset \mathbb{R}$ and consider the Hilbert space

$$H = L^2(a, b; w(x)dx).$$

Assume furthermore that

$$\int_a^b |x|^n w(x)dx < \infty$$

for all n (this follows from $w \in L^1(a,b)$ if the interval is bounded.)

Define $P_n(x)$ to be the unique polynomial of degree n whose highest order term is x^n, such that $P_n \perp x^j$ for all $j < n$. The polynomials thus defined are the "orthogonal polynomials on (a,b) with weight $w(x)$." They form an orthogonal system.

Exercise 127. Suppose (a,b) is a bounded interval. Prove that $\{P_n : n \in \mathbb{N}\}$ forms a complete orthogonal system. (You could use the fact that continuous functions are dense in $L^2(a,b,w(x)dx)$ combined with the Stone-Weierstrass theorem to show that polynomials are also dense.)

Exercise 128. (Legendre Polynomials)
Compute P_0, P_1, P_2, P_3 for $(a,b) = (-1,1)$ and $w(x) = 1$.
Show that

$$P_n(x) = \frac{n!}{(2n)!}\left(\frac{d}{dx}\right)^n (1-x^2)^n.$$

Hint: Use integration by parts to prove that the P_n given by this formula are orthogonal. then compute the coefficient of x^n in P_n.

Theorem 74. *Let $P_n(x)$ be the sequence of orthogonal polynomials in $L^2(a,b;w(x)dx)$. Then $P_n(x)$ has n zeroes in the interval (a,b) so that P_n can be written as*

$$P_n(x) = A_n(x-x_1)(x-x_2)\dots(x-x_n)$$

for certain $a < x_1 < x_2 < \dots < x_n < b$ and $A_n \neq 0$.

Proof. If P_n has only $k < n$ sign changes, say at $x_1, \dots, x_k \in (a,b)$, then we consider $Q(x) = (x-x_1)\dots(x-x_k)$. The degree of Q is $k < n$ so $Q \perp P_n$, but $Q(x)P_n(x)$ does not change sign so that

$$\int_a^b Q(x)P_n(x)w(x)dx \neq 0.$$

This contradicts $Q \perp P_n$. Therefore P_n has n sign changes, and since P_n is a polynomial of degree n these sign changes must be simple zeroes, i.e.

$$P_n(x) = A_n(x-x_1)(x-x_2)\dots(x-x_n).$$

□

Exercise 129. Show that there exist constants $b_n c_n \in \mathbb{R}$ such that

$$xP_n(x) = b_n P_{n+1}(x) + c_n P_n(x) + b_n P_{n-1}(x).$$

Hint: $(xP_{n+1}, P_m) = (P_{n+1}, xP_m)$.

70. The Spectral Theorem for Symmetric Compact Operators

One of the main theorems from linear algebra states that every symmetric $n \times n$ matrix T has an orthonormal basis of eigenvectors $v_1, \dots, v_n \in \mathbb{R}^n$ with real eigenvalues. Thus $\{v_1, \dots, v_n\}$ is a complete orthonormal set in $\mathbb{R}^n$ and $Tv_i = \lambda_i v_i$. With respect to the basis $\{v_1, \dots, v_n\}$ the matrix of T is diagonal.

Similar theorems exist for operators in Banach spaces. The most successful and complete is the spectral theorem for self-adjoint operators in a Hilbert space. In this section we will state and prove a version for compact symmetric operators.

Definition. Let H be a separable Hilbert space, and let $T : H \to H$ be an operator. By definition T is *symmetric* if $(Tf, g) = (f, Tg)$ for all $f, g \in H$.

Exercise 130. *Symmetric Hilbert Schmidt operators.* An integral operator defined by a kernel $T(x, y)$, with $T \in L^2(\Omega \times \Omega)$ is symmetric if and only if $T(x, y) \equiv T(y, x)$ (almost everywhere).

Exercise 131. Let $\Omega = (0, 1)$, and let $h : \Omega \to \mathbb{C}$ be a complex valued measurable function. Show that the operator $Tf(x) = h(x)f(x)$ is symmetric on $L^2(\Omega)$ iff f is real valued.

Exercise 132. Let T be the operator $Tf(x) = xf(x)$ on $L^2(\Omega)$, $\Omega = (0, 1)$. Does this operator have eigenvalues, i.e. are there $\lambda \in \mathbb{C}$ and $f \in L^2(\Omega)$ with $Tf = \lambda f$?

Theorem 75 (Spectral Theorem)**.** *Let H be a separable Hilbert space. If $T : H \to H$ is compact and symmetric then there is a complete orthonormal basis for H consisting entirely of eigenvectors of T.*

Proof of the spectral theorem

We divide the proof into a sequence of lemmas.

Lemma 76 (The largest and smallest eigenvalues.)**.** *If $T : H \to H$ is symmetric and compact then*

$$\lambda(T) = \sup\{(Tx, x) : \|x\| \leq 1\}, \text{ and } \mu(T) = \inf\{(Tx, x) : \|x\| \leq 1\}$$

both are attained.

If $\lambda(T) > 0$ and $(Tx, x) = \lambda(T)$ with $\|x\| = 1$ then x is an eigenvector of T with eigenvalue $\lambda(T)$.

If $\mu(T) < 0$ and $(Tx, x) = \mu(T)$ with $\|x\| = 1$ then x is an eigenvector of T with eigenvalue $\mu(T)$.

Proof. If $\lambda(T) = 0$ then the maximum is attained at $x = 0$.

Assume $\lambda(T) > 0$, and let $x_i \in H$ be a sequence with $\|x_i\| \leq 1$ and $\lim_{i\to\infty}(Tx_i, x_i) = \lambda(T)$.

Since H is a Hilbert space H is reflexive. We may therefore use the Banach-Alaoglu theorem to extract a weakly convergent subsequence of the x_i, which we denote by x_i again.

Since T is compact we may extract a further subsequence for which Tx_i converges in the norm of H. We again denote this second subsequence by x_i.

Thus we have $x_i \rightharpoonup x_*$ and $Tx_i \to y$. For an arbitrary $z \in H$ we have

$$(y, z) = \lim_{i\to\infty}(Tx_i, z) = \lim_{i\to\infty}(x_i, Tz) = (x_*, Tz) = (Tx_*, z).$$

This implies that $y = Tx_*$.

We now claim that $(x_*, Tx_*) = \lambda(T)$. To see this we observe that

$$\begin{aligned}\lambda(T) &= \lim_{i\to\infty}(x_i, Tx_i) \\ &= \lim_{i\to\infty}\{(x_i, y) + (x_i, Tx_i - y)\}\end{aligned}$$

Here the first term converges to $(x_*, y) = (x_*, Tx_*)$ while the second term is bounded by

$$|(x_i, Tx_i - y)| \leq \|x_i\| \, \|Tx_i - y\| \leq \|Tx_i - y\| \to 0.$$

We therefore have $\lambda(T) = (x_*, Tx_*)$ as claimed: The quantity (x, Tx) attains a maximum on $\{x : \|x\| \leq 1\}$.

A similar argument shows that (x, Tx) also attains a minimum.

It remains to show that if $\lambda(T) > 0$ then the maximum is attained at an eigenvector of T.

First we observe that when $\lambda(T) > 0$ the maximizing vector x is a unit vector. Indeed, if $\|x\| < 1$, then there is a $c > 1$ with $\|cx\| = 1$, and one would have $\lambda(T) \geq (cx, T(cx)) = c^2(x, Tx) = c^2\lambda(T) > \lambda(T)$.

Next, let $z \in H$ be any unit vector perpendicular to x. Consider

$$x(t) = (\cos t)x + (\sin t)z, \quad t \in \mathbb{R}$$

For any $t \in \mathbb{R}$ $x(t)$ is a unit vector and hence $(x(t), Tx(t)) \leq \lambda(T)$ with equality for $t = 0$. Thus the real valued function $f(t) = (x(t), Tx(t))$ attains a maximum at $t = 0$ and we must have $f'(0) = 0$. Now compute

$$\begin{aligned} f'(0) &= \frac{d}{dt}\bigg|_{t=0} \left((x, Tx)\cos^2 t + 2(Tx, z)\sin t \cos t + (z, Tz)\sin^2 t \right) \\ &= 2(Tx, z) \end{aligned}$$

Thus $Tx \perp z$ for all $z \perp x$: this can only happen if Tx is a multiple of x. Since $(Tx, x) = \lambda(T)$ it follows that $Tx = \lambda(T)x$.

□

Lemma 77 (Invariant orthogonal splittings.)**.** *Let $V \subset H$ be a linear subspace which is invariant under the symmetric operator T, i.e. $T(V) \subset V$. Then the orthogonal complement $V^\perp = \{x \in H : x \perp V\}$ is also invariant under T.*

Proof. If V is invariant then we consider any $x \in V^\perp$ and note that for all $v \in V$ one has $Tv \in V$ by assumption and hence,

$$(v, Tx) = (Tv, x) = 0.$$

Thus $Tx \perp V$. □

Construction of all eigenvectors and eigenvalues. Define numbers

$$\lambda_1 \geq \lambda_2 \geq \cdots > 0 > \cdots \geq \mu_2 \geq \mu_1,$$

and vectors $x_1, x_2, \cdots, y_1, y_2, \cdots$ by the following inductive process.

Let $\lambda_1 = \lambda(T)$, and let x_1 be a corresponding eigenvector; similarly, let $\mu_1 = \mu(T)$, and let y_1 be a corresponding eigenvector.

Given $x_1, \cdots, x_{n-1}$ and $y_1, \cdots, y_{n-1}$ let H_{n-1} be the orthogonal complement of $x_1, \cdots, x_{n-1}$ and $y_1, \cdots, y_{n-1}$. Since the x_i and y_i are eigenvectors of T, the space H_{n-1} is invariant under T, and we can define

$$\begin{aligned} \lambda_n &= \sup\{(Tx, x) : \|x\| \leq 1, x \in H_{n-1}\} = \lambda(T|H_{n-1}), \\ \mu_n &= \inf\{(Ty, y) : \|y\| \leq 1, y \in H_{n-1}\} = \mu(T|H_{n-1}). \end{aligned}$$

If $\lambda_n > 0$ we let x_n be a point in the unit ball of H_{n-1} where the sup is attained, and likewise, if $\mu_n < 0$ we choose a y_n which minimizes (Ty, y) over the unit

ball in H_{n-1}. In either case x_n and y_n are eigenvectors with eigenvalues λ_n, μ_n respectively.

If for some n the λ_n turns out to vanish, one ends the sequence of λ_i's and x_i's at $i = n-1$, and similar measures are to be taken if μ_n happens to be zero.

Lemma 78. $\lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} \mu_n = 0$

Proof. The same arguments apply to both λ_n and μ_n. We only deal with the former.

The λ_n form a decreasing sequence so $\Lambda = \lim_{n\to\infty} \lambda_n \geq 0$ exists. Suppose $\Lambda > 0$. Since the sequence x_n is bounded and T is compact, the image Tx_n must have a convergent subsequence. But then convergence of λ_n and Tx_n implies that $x_n = \lambda_n^{-1} T x_n$ also converges. On the other hand the x_n are orthogonal unit vectors so that $\|x_n - x_m\| = \sqrt{2}$ for all $n \neq m$. There can be no convergent subsequence! □

Let H_∞ be the intersection of all H_n. On H_n one has

$$\mu_n \|x\|^2 \leq (Tx, x) \leq \lambda_n \|x\|^2$$

so on the intersection H_∞ one has $(Tx, x) \equiv 0$. Hence

$$0 = (T(x+y), (x+y)) = (Tx, x) + (Ty, y) + 2\,(Tx, y) = 2\,(Tx, y)$$

for all $x, y \in H_\infty$, which means that T vanishes on H_∞ (take $y = Tx$.)

If one now chooses an arbitrary complete orthonormal basis $\{z_i\}$ for H_∞, then $\{x_i\} \cup \{y_i\} \cup \{z_i\}$ is a complete orthonormal basis for H which diagonalizes T. The proof of the spectral theorem is complete.

What happened in this proof? The following problem reveals a nice fact about the "eigenvalues" of a symmetric matrix which doesn't show up in most introductions to matrix algebra.

Exercise 133. Solve the following (3rd semester, honors) Calculus problem: Let $A = (a_{ij})$ be a symmetric $n \times n$ matrix. Find the maxima and minima of

$$f(x_1, \ldots, x_n) = \tfrac{1}{2} \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

on the unit sphere, i.e. find the extrema of f subject to the constraint $x_1^2 + \ldots + x_n^2 = 1$.

The following description of eigenvalues turns out to be of practical use in computing, say, the eigenvalues of the Laplacian on some bounded domain in $\mathbb{R}^n$. Here one can interpret "computing" in the sense of numerical computation, involving a fast computer to obtain $\lambda_1, \ldots, \lambda_{25}$ in several digits, but also in the sense of estimating the λ_n by hand, e.g. to obtain information about the growth rate of the eigenvalues λ_n as $n \to \infty$.

Exercise 134. *Rayleigh's minimax characterization of the eigenvalues.*
Let $G_k(H)$ be the set of k-dimensional subspaces of H. Then

$$\lambda_n = \sup_{V \in G_k(H)} \inf_{x \in V} \frac{(Tx, x)}{\|x\|^2},$$

$$\mu_n = \inf_{V \in G_k(H)} \sup_{x \in V} \frac{(Tx, x)}{\|x\|^2}.$$

71. Eigenfunctions of the Laplacian

This section is devoted to a proof of

Theorem 79. *Let $\Omega \subset \mathbb{R}^n$ be a bounded domain. Then there is a complete orthonormal basis $\{\phi_n : n \in \mathbb{N}\}$ for $L^2(\Omega)$ consisting of eigenfunctions of the Laplace operator, i.e. the ϕ_n belong to $W_o^{1,2}(\Omega)$ and*

$$-\Delta\phi_n = \lambda_n \phi_n \qquad (\text{in } \mathcal{D}'(\Omega))$$

for certain constants

$$\lambda_n > \frac{1}{4}|\Omega|^{-2/n}.$$

We will prove this theorem by applying the spectral theorem to Green's operator.

Recall that Green's operator was defined in §63 by saying that $Gf = u$ if $f \in L^2(\Omega)$, if $u \in W_o^{1,2}(\Omega)$ and if u satisfies

$$-\Delta u = f$$

in the sense of distributions. It was shown in §63 that G is a bounded linear and compact operator.

Lemma 80. *Green's operator G is symmetric and strictly positive definite, i.e. $(Gf, f) > 0$ for all $f \neq 0 \in L^2(\Omega)$.*

Proof. We must show that $(Gf, g) = (f, Gg)$ for all $f, g \in L^2(\Omega)$.

From the construction of G we know that $Gf = u$ if u minimizes

$$Q_f(u) = \int_\Omega \left\{ \tfrac{1}{2}|\nabla u(x)|^2 - u(x)f(x) \right\} \, dx$$

over all $u \in W_o^{1,2}(\Omega)$.

Similarly, if $Gg = v$ then v minimizes

$$Q_g(v) = \int_\Omega \left\{ \tfrac{1}{2}|\nabla v(x)|^2 - v(x)g(x) \right\} \, dx$$

over all $v \in W_o^{1,2}(\Omega)$.

Hence $Q_f(u + tv)$, as a function of $t \in \mathbb{R}$, attains a minimum at $t = 0$. Hence

$$\begin{aligned} 0 &= \left. \frac{dQ_f(u+tv)}{dt} \right|_{t=0} \\ &= \left. \frac{d}{dt} \right|_{t=0} \left(\frac{1}{2} \int |\nabla u|^2 dx + t \int \nabla u \cdot \nabla v dx \right. \\ &\qquad \left. + \frac{t^2}{2} \int |\nabla v|^2 dx - \int uf dx - t \int vf dx \right) \\ &= \int \nabla u \cdot \nabla v dx - \int vf dx. \end{aligned}$$

Thus

$$\int \nabla u \cdot \nabla v dx = \int_\Omega v(x) f(x) \, dx.$$

Applying the same argument to $Q_g(v + tu)$ we also get

$$\int \nabla u \cdot \nabla v dx = \int_\Omega u(x) g(x) \, dx.$$

Thus we have

$$(Gf, g) = (u, g) = \int \nabla u \cdot \nabla v dx = (v, f) = (Gg, f).$$

This shows that G is symmetric. By setting $u = v$ and $f = g$ we get

$$(Gf, f) = \int |\nabla u|^2 \, dx \geq 0$$

with equality only for $u = 0$ and hence $f = -\Delta u = 0$. □

Since G is a compact operator we may apply the spectral theorem: There exists an orthonormal family $\phi_n \in L^2(\Omega)$ of eigenfunctions of G,

$$G\phi_n = \gamma_n \phi_n.$$

Lemma 81. *$0 < \gamma_n \leq 4|\Omega|^{2/n}$ for all n.*

Proof. Since G is strictly positive definite we have

$$\gamma_n = \gamma_n(\phi_n, \phi_n) = (\phi_n, G\phi_n) > 0$$

for all n. From $\|G\| \leq$ we get

$$\gamma_n = \gamma_n \|\phi_n\| = \|\gamma_n \phi_n\| = \|G\phi_n\| \leq 4|\Omega|^{2/n} \|\phi_n\| = 4|\Omega|^{2/n}.$$

□

From $G\phi_n = \gamma_n \phi_n$ and the definition of G it follows that the ϕ_n satisfy

$$\gamma_n \left(-\Delta \phi_n\right) = \phi_n,$$

in the sense of distributions. Hence

$$-\Delta \phi_n = \lambda_n \phi_n$$

where $\lambda_n = \frac{1}{\gamma_n} \geq \frac{1}{4}|\Omega|^{-2/n}$.

An example: the one dimensional case.

If $\Omega = (0, 1)$, then we have shown that there is a complete orthonormal system of functions $\phi_k(x)$ which satisfies

$$\phi_k''(x) + \lambda_k \phi_k(x) = 0, \quad (0 < x < 1)$$
$$\phi_k(0) = \phi_k(1) = 0$$

Here the first line is meant in the sense of distributions, and the second should be interpreted as $\phi_k \in W_o^{1,2}(\Omega)$. However, we have seen that in one dimension $W_o^{1,2}$ functions are continuous, and that they vanish on $\partial\Omega = \{0, 1\}$.

Starting from $\phi \in C([0, 1])$ one shows by induction that the differential equation $\phi'' = -\lambda\phi$ implies that $\phi \in C^\infty([0, 1])$. We can then use our "cookbook differential equation knowledge" to find the ϕ_k. They are:

$$\phi_k(x) = \sqrt{2}\sin(k\pi x), \quad \lambda_k = \left(k\pi\right)^2, \quad (k = 1, 2, 3, \dots).$$

What are the eigenvalues?

"And now for a message from our sponsor ... "
(M.Spivak, Differential geometry, vol.5)

Historically the eigenfunctions of the Laplacian are attached to the heat and wave equations. The *wave equation* is the PDE

$$\frac{\partial^2 u}{\partial t^2} = \Delta u, \quad x \in \Omega, t \geq 0. \tag{38}$$

It describes the following physical situation (and many others): if $\Omega \subset \mathbb{R}^2$ is bounded, then one can imagine a membrane in the shape of Ω whose boundary $\partial\Omega$ is kept fixed, but which is allowed to vibrate. If one represents the vertical deviation at $x \in \Omega$ and time $t \geq 0$ by a function $u(x,t)$ then, assuming the deviations are small, and the membrane is of uniform thickness, etc. one arrives at (38).

If ϕ_n is an eigenfunction of Δ with eigenvalue λ_n, then direct substitution shows that

$$S_n(x,t) \stackrel{\text{def}}{=} \sin(\sqrt{\lambda_n}t)\phi_n(x), \text{ and } C_n(x,t) \stackrel{\text{def}}{=} \cos(\sqrt{\lambda_n}t)\phi_n(x)$$

are solutions to (38) which satisfy the boundary condition

$$u(x,t) = 0 \text{ for } x \in \partial\Omega \text{ and } t \geq 0. \tag{39}$$

Since the wave equation is linear any linear combination of the V_n and U_n is again a solution. Thus one arrives at the "general solution"

$$u(x,t) = \sum_{k=1}^{\infty} \Big\{ A_k \sin(\sqrt{\lambda_k}t)\phi_k(x) + B_k \cos(\sqrt{\lambda_k}t)\phi_k(x) \Big\}$$

where the coefficients A_k, B_k are determined by the initial position and velocities

$$u(x,0) = \sum_{k=1}^{\infty} B_k \phi_k(x)$$

$$u_t(x,0) = \sum_{k=1}^{\infty} A_k \sqrt{\lambda_k}\phi_k(x).$$

In particular, this shows that any solution is a superposition of harmonic (sinusoidal) vibrations with frequencies $\omega_k = \sqrt{\lambda_k}$.

For the heat equation one has the same story. The heat (or diffusion) equation is

$$\frac{\partial u}{\partial t} = \Delta u, \quad x \in \Omega, t \geq 0. \tag{40}$$

If one thinks of Ω as a solid whose temperature $u(x,t)$ at point x and time t is not constant, then under various assumptions it follows that the temperature must obey (40).

Assume furthermore that the boundary $\partial\Omega$ of the solid Ω is kept at a constant temperature $u = 0$ (e.g. the solid is submerged in melting ice) and it follows that the temperature must satisfy the boundary condition (39).

As with the heat equation one can find simple solutions of the form

$$H_n(x,t) = e^{-\lambda_n t}\phi_n(x)$$

and then exploit linearity of the equation and boundary condition to obtain the "general solution"

$$u(x,t) = \sum_{k=1}^{\infty} C_k e^{-\lambda_n t} \phi_n(x).$$

Again, the constants C_k are to be determined from the initial temperature by expanding $u(x,0)$ with respect to the complete orthonormal system $\{\phi_k(x)\}_{k\in\mathbb{N}}$, i.e.

$$u(x,0) = \sum_{k=1}^{\infty} C_k \phi_k(x) \Rightarrow C_k = (u(\cdot,0),\phi_k)_{L^2(\Omega)}.$$

It was the one dimensional version of this problem (heat conduction in an interval with periodic boundary conditions) which led Fourier to study the series named after him.

For an entertaining article (which spawned several research papers in the years since then) about the eigenvalues of the Laplacian see M.Kac's *"Can you hear the Shape of a Drum?"* American Mathematical Monthly, 1966, and a more recent follow-up by M. H. Protter, *"Can one hear the shape of a drum? revisited"* SIAM Review **29** (1987) pp.185–197.

The Fourier transform

Fourier series; The Fourier transform; the Inversion Formula; Tempered distributions; Plancherel's Formula; Fourier multipliers; Elliptic regularity;

Textbooks to look at: Rudin's Real and Complex book [4] has chapters on both Fourier series and the Fourier transform. In [5] he also treats the Fourier transform from the point of view of tempered distributions. A more specialized book on Fourier analysis only is Katznelson's [10].

72. Fourier series

In dealing with Fourier series we consider periodic functions with period 2π. To fix notation we let $\mathbb{T}$ be the unit circle in $\mathbb{C}$. Points on $\mathbb{T}$ are of the form $e^{ix} = \cos x + i \sin x$, and functions $f : \mathbb{T} \to \mathbb{R}$ can be regarded as 2π periodic functions of $x \in \mathbb{R}$. When integrating such functions one has

$$\int_{\mathbb{T}} f(x)dx = \int_0^{2\pi} f(x)dx = \int_a^{2\pi+a} f(x)dx = \int_{-\pi}^{\pi} f(x)dx.$$

For instance the convolution $f * g$ of $f, g \in L^1(\mathbb{T})$ is given by

$$\begin{aligned} f * g(x) &= \int_{\mathbb{T}} f(x-y)g(y)\, dy \\ &= \int_0^{2\pi} f(x-y)g(y)\, dy \\ &= \int_{x-2\pi}^{x} f(x-y)g(y)\, dy \qquad\qquad (y := x - z) \\ &= \int_0^{2\pi} f(z)g(x-z)\, dy \\ &= g * f(x). \end{aligned}$$

We will use these identities freely.

Exercise 135. Show that for $f, g \in L^1(\mathbb{T})$ the convolution $f * g$ again belongs to $L^1(\mathbb{T})$, and that one has

$$\|f * g\|_{L^1(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})}\|g\|_{L^1(\mathbb{T})}.$$

Prove that for $f \in L^1(\mathbb{T})$ and $g \in L^p(\mathbb{T})$ (where $1 \leq p \leq \infty$) one has $f * g \in L^p(\mathbb{T})$ with

$$\|f * g\|_{L^p(\mathbb{T})} \leq \|f\|_{L^1(\mathbb{T})}\|g\|_{L^p(\mathbb{T})}.$$

Exercise 136. Let $h \in L^1(\mathbb{T})$ be given, and consider the bounded operator on $L^2(\mathbb{T})$ defined by convolution with h, i.e.

$$Tf(x) = h * f(x) = \int_0^{2\pi} h(y)f(x-y)dy$$

Show that the functions $e_n(x) = \frac{1}{\sqrt{2\pi}}e^{2\pi nx}$ are eigenfunctions of T. What are the corresponding eigenvalues?

For $f \in L^1(\mathbb{T})$ we define the Fourier coefficients to be

$$\hat{f}_n = \frac{1}{2\pi}\int_{\mathbb{T}} e^{-inx}f(x)\ dx$$

The sequence of Fourier coefficients is bounded by

$$\sup|\hat{f}_n| \leq \frac{1}{2\pi}\|f\|_{L^1(\mathbb{T})}.$$

The map $f \mapsto \{\hat{f}_n : n \in \mathbb{Z}\}$ is therefore a bounded linear map from $L^1(\mathbb{T})$ to $\ell^\infty(\mathbb{Z})$.

We have seen that for $f \in L^2$ the partial sums

$$s_N f(x) = \sum_{n=-N}^{N} \hat{f}_n e^{inx}$$

converge in L^2 to f. It is natural to ask if $s_N f(x)$ converges pointwise or in any other sense to f: the answer to this question is surprisingly complicated and has a very long history.

There are *many* convergence theorems. In these notes I will only prove one of them (Theorem 84 below). In Katznelson's [10, chapter I&II] you can find much more on the convergence and divergence of Fourier series.

72.1. The Dirichlet kernel

Theorem 82. *The partial sums $s_N f$ of a function $f \in L^1(\mathbb{T})$ admit the following explicit representation,*

$$s_N f(x) = \int_{\mathbb{T}} D_N(x-y)f(y)\ dy \tag{41}$$

where

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

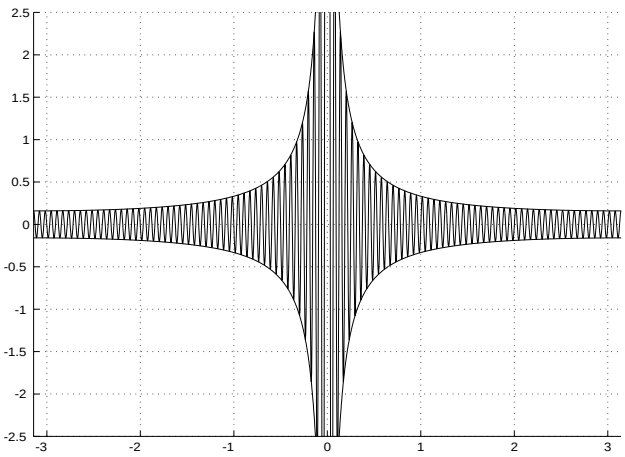


FIGURE 2. The Dirichlet kernel for $N = 100$.

Proof. One has

$$\begin{aligned} s_N f(x) &= \sum_{k=-N}^{N} \hat{f}_k e^{2k\pi ix} \\ &= \sum_{k=-N}^{N} \frac{e^{2k\pi ix}}{2k\pi} \int_{\mathbb{T}} f(y) e^{-2k\pi iy}\, dy \\ &= \int_{\mathbb{T}} \left\{ \sum_{k=-N}^{N} \frac{e^{2k\pi i(x-y)}}{2\pi} \right\} f(y)\, dy \\ &= \int_{x_0-\pi}^{x_0+\pi} D_N(x-y) f(y)\, dy \end{aligned}$$

provided

$$D_N(t) = \frac{1}{2\pi} \sum_{k=-N}^{N} e^{2k\pi it}.$$

Summation of the geometric series in $e^{2\pi it}$ gives the other expression for $D_N(t)$. □

The highly oscillatory function D_N is called the *Dirichlet kernel.* (See figure 2.) A basic property of D_N is

$$\int_{\mathbb{T}} D_N(x)\, dx = 1. \tag{42}$$

One verifies this by integrating $D_N(x) = (2\pi)^{-1} \sum e^{2k\pi ix}$ term by term.

To prove convergence of the partial sums one must use the oscillatory nature of D_N, so we recall

Lemma 83 (Riemann-Lebesgue)**.** *If $f \in L^1(\mathbb{T})$ then for any $x_0 \in \mathbb{R}$*

$$\lim_{A\to\infty} \int_0^{2\pi} \sin A(x-x_0) f(x)\, dx = 0.$$

You have proved this in several exercises in these notes for the case $f \in L^1(\mathbb{R})$. The case of periodic functions follows immediately: given a periodic function $f \in L^1(\mathbb{T})$ apply the $L^1(\mathbb{R})$ version of the Riemann Lebesgue Lemma to the function

$$\tilde{f}(x) \stackrel{\text{def}}{=} \begin{cases} f(x) & x \in (0, 2\pi) \\ 0 & \text{elsewhere.} \end{cases}$$

By definition a continuous function $f : \mathbb{T} \to \mathbb{C}$ is said to be *Dini-continuous* at a point $x_0 \in \mathbb{T}$ if

$$\int_{\mathbb{T}} \frac{|f(x) - f(x_0)|}{|x - x_0|}\, dx < \infty$$

Exercise 137. Verify that any Hölder continuous function is everywhere Dini continuous.

Theorem 84. *If f is Dini-continuous at x_0 then*

$$\lim_{N\to\infty} s_N f(x_0) = f(x_0).$$

Proof. One has, using (42),

$$\begin{aligned} s_N f(x_0) - f(x_0) &= \int_{\mathbb{T}} D_N(x - y) f(y)\, dy - f(x_0) \\ &= \int_{\mathbb{T}} D_N(x - y)\{f(y) - f(x_0)\}\, dy \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} \sin(N + \tfrac{1}{2})(x_0 - y) \frac{f(y) - f(x_0)}{\sin \frac{1}{2}(y - x_0)}\, dy \\ &= \frac{1}{2\pi} \int_{\mathbb{T}} \sin(N + \tfrac{1}{2})(x_0 - y) g(y)\, dy \end{aligned}$$

where

$$g(y) \stackrel{\text{def}}{=} \frac{f(y) - f(x_0)}{\sin \frac{1}{2}(y - x_0)} = \frac{f(y) - f(x_0)}{y - x_0} \; \frac{y - x_0}{\sin \frac{1}{2}(y - x_0)}$$

Since $\frac{t}{\sin(t/2)} \leq \pi$ for $|t| \leq \pi$ we have

$$\left| \frac{y - x_0}{\sin \frac{1}{2}(y - x_0)} \right| \leq \pi$$

for $x_0 - \pi < y < x_0 + \pi$. Thus $g(y)$ is integrable on the interval $x_0 - \pi < y < x_0 + \pi$, and the Riemann-Lebesgue Lemma implies that

$$\lim_{N\to\infty} s_N f(x_0) - f(x_0) = \lim_{N\to\infty} \frac{1}{2\pi} \int_{\mathbb{T}} \sin(N + \tfrac{1}{2})(x_0 - y) g(y)\, dy = 0.$$

□

Exercise 138. For which $x \in \mathbb{T}$ does this theorem apply to the functions whose Fourier series you computed in problems 122, 123 and 124?

72.2. A divergent Fourier Series

Theorem 85. *A continuous function $g \in C(\mathbb{T})$ exists whose Fourier series diverges at $x = 0$.*

Proof. An explicit construction can be given, but in this case Functional Analysis (Baire's theorem in the form of the Banach-Steinhaus theorem) provides a shortcut.

Consider the functional

$$\lambda_N(f) \stackrel{\text{def}}{=} s_N f(0) = \frac{1}{2\pi} \sum_{k=-N}^{N} \int_{\mathbb{T}} e^{ikx} f(x)\, dx.$$

It is clearly a bounded functional on $C(\mathbb{T})$. A direct calculation show that for

$$f_N(x) = \sin(N + \tfrac{1}{2})x, \qquad 0 < x < 2\pi.$$

one has

$$\lambda_N(f_N) = \frac{2}{\pi} \left\{ 1 + \frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{4N+1} \right\}.$$

(The 2π periodic extension of f_N is continuous, but not differentiable at $x = 0$!)

Since $\|f_N\|_\infty = 1$, this implies that

$$\|\lambda_N\|_{C(\mathbb{T})^*} \geq \lambda_N(f_N) \to \infty, \text{ as } N \to \infty.$$

If $\sup_N |\lambda_N(g)|$ were finite for every $g \in C(\mathbb{T})$ then the Banach-Steinhaus theorem would imply that the functionals λ_N are bounded, i.e. $\sup_N \|\lambda_N\| < \infty$. We have just shown that this is not the case so we conclude that there exists a $g \in C(\mathbb{T})$ for which the sequence $\{\lambda_N g \in C(\mathbb{T}) : N \in \mathbb{N}\}$ is unbounded. The Fourier series of this g can therefore not converge at $x = 0$. □

With much more work one can prove the following.

Theorem 86 (Kolmogorov)**.** *A function $f \in L^1(\mathbb{T})$ exists whose Fourier series diverges everywhere.*

Theorem 87 (L.Carleson, 1965)**.** *If $f \in L^2(\mathbb{T})$ then the Fourier series of f converges almost everywhere to $f(x)$.*

Theorem 88 (Y.Katznelson & J.-P.Kahane, 1966)**.** *For every set $E \subset \mathbb{T}$ of zero measure there is a continuous function f whose Fourier series diverges precisely on E.*

72.3. The Fejér kernel

By explicit computation one finds

$$\begin{aligned} \sigma_N f(x) &\stackrel{\text{def}}{=} \frac{s_0 f(x) + s_1 f(x) + \ldots + s_N f(x)}{N+1} \\ &= \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N+1}\right) \hat{f}_k e^{ikx} \\ &= \int_{\mathbb{T}} K_N(x-y) f(y)\, dy \end{aligned}$$

where

$$K_N(x) = \frac{1}{N+1} \left(\frac{\sin \frac{N+1}{2} x}{\sin \frac{1}{2} x} \right)^2$$

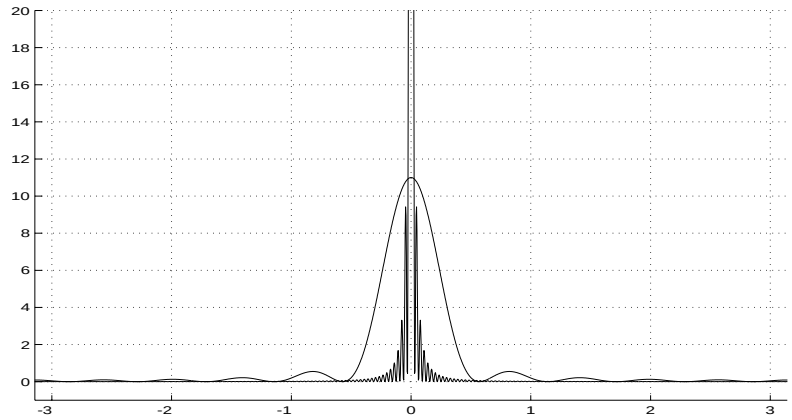


FIGURE 3. The Fejer kernels for $N = 10$ and $N = 200$.

is the *Fejér kernel.* This kernel is much better behaved than the Dirichlet kernel (see figure 72.2), and the average partial sums $\sigma_N f(x)$ converge more often than the partial sums $s_N f(x)$ themselves.

Theorem 89. *For any $f \in C(\mathbb{T})$ the average partial sums $\sigma_N f$ of the Fourier series of f converge uniformly to f.*

73. The Fourier Integral

For $f \in L^1(\mathbb{R}^n)$ we define the Fourier transform of f to be

$$\mathcal{F}f(x) = \int_{\mathbb{R}^n} e^{-2\pi i x\xi} f(x)\, dx.$$

The inverse Fourier transform of a $g \in L^1(\mathbb{R}^n)$ is

$$\mathcal{F}^* g(\xi) = \int_{\mathbb{R}^n} e^{2\pi i x\xi} g(\xi)\, d\xi.$$

Other notation which is used widely is

$$\mathcal{F}f(\xi) = \hat{f}(\xi), \quad \mathcal{F}^* g(x) = \check{g}(x).$$

Theorem 90 (Basic Properties of $\mathcal{F}$)**.**

1. *If $f \in L^1(\mathbb{R}^n)$ then $\mathcal{F}f \in C(\mathbb{R}^n)$ and $\lim_{|\xi|\to\infty} \mathcal{F}f(\xi) = 0$.*
2. *For $f, g \in L^1$ one has*
$$\mathcal{F}(f * g) = (\mathcal{F}f)(\mathcal{F}g)$$
3. *If $f, D_k f \in L^1(\mathbb{R}^n)$ then*
$$\mathcal{F}(D_k f)(\xi) = 2\pi i \xi_k \mathcal{F}f(\xi).$$
4. *If $(1 + |x|)f \in L^1(\mathbb{R}^n)$ then $\mathcal{F}f \in C^1(\mathbb{R}^n)$ and*
$$D_k \mathcal{F}f = \mathcal{F}(-2\pi i x_k f).$$

The proofs are left as an exercise. (But see Rudin [5] if you get stuck.)

74. The Inversion Theorem

There are several versions of the inversion theorem, all of which guarantee

$$\mathcal{F}^*(\mathcal{F}(f)) = f$$

under varying hypotheses on f, and under various interpretations of "=" (pointwise, almost everywhere, in the sense of (tempered) distributions, etc.)

Theorem 91. (a) *If $f \in L^1(\mathbb{R}^n)$ and $\mathcal{F}f \in L^1(\mathbb{R}^n)$ then $f(x) = (\mathcal{F}^*\mathcal{F}f)(x)$ for almost all $x \in \mathbb{R}^n$.*

(b) *If $f \in L^1(\mathbb{R}^n)$ and $\mathcal{F}^*f \in L^1(\mathbb{R}^n)$ then $f(x) = (\mathcal{F}\mathcal{F}^*f)(x)$ for almost all $x \in \mathbb{R}^n$.*

Note that $\mathcal{F}f \in L^1$ implies that $\mathcal{F}^*\mathcal{F}f \in C_0(\mathbb{R}^n)$ so that this theorem says that f is almost everywhere equal to a continuous function.

Proof. The second statement is obtained from the first by changing i to $-i$ in the definition of the Fourier transform. Below we will prove (a); by changing all i's to $-i$'s you get a proof of (b).

For the proof you must choose a bounded continuous function $m : \mathbb{R}^n \to \mathbb{R}$ whose Inverse Fourier transform $M = \mathcal{F}^*m$ you know, and for which m and M are both in $L^1(\mathbb{R}^n)$ with $m(0) = 1$ and

$$\int_{\mathbb{R}^n} M(x)\, dx = 1. \tag{43}$$

Here are some m's you could choose:

$m(x)$	$M(\xi)$
$e^{-\lvert x\rvert^2}$	$\pi^{n/2}e^{-\pi^2\lvert\xi\rvert^2}$
$\prod_{j=1}^n (1-\lvert x_j\rvert)_+$	$\prod_{j=1}^n \left(\dfrac{\sin \pi\xi_j}{\pi\xi_j}\right)^2$
$e^{-\lvert x_1\rvert-\lvert x_2\rvert-\cdots-\lvert x_n\rvert}$	$\dfrac{2^n}{(1+4\pi^2\xi_1^2)\cdots(1+4\pi^2\xi_n^2)}$

Denote

$$m_\epsilon(\xi) = m(\epsilon\xi), \qquad M_\epsilon(x) = \epsilon^{-n} M\left(\tfrac{x}{\epsilon}\right).$$

On then establishes the following identity:

$$\mathcal{F}^*(m_\epsilon(\xi)(\mathcal{F}f)(\xi)) = M_\epsilon * f. \tag{44}$$

We now let $\epsilon \searrow 0$ on both sides. On the left we get

$$\mathcal{F}^*(m_\epsilon \cdot \mathcal{F}f)(x) \to (\mathcal{F}^*\mathcal{F})f(x) \text{ uniformly in } x \in \mathbb{R}^n.$$

Indeed, $m(\epsilon\xi)$ is uniformly bounded, and converges pointwise to $m(0)$ as $\epsilon \searrow 0$, while $\mathcal{F}^*f \in L^1$. Hence the Dominated Convergence theorem applies to

$$\mathcal{F}^*(m_\epsilon \cdot \mathcal{F}f)(x) = \int_{\mathbb{R}^n} e^{2\pi i x\xi} m(\epsilon\xi)\mathcal{F}^*f(\xi)\, d\xi.$$

On the right in (44) we have

$$\begin{aligned} M_\epsilon * f &= \int_{\mathbb{R}^n} \epsilon^{-n} M\left(\tfrac{x-y}{\epsilon}\right) f(y)\, dy \qquad\qquad (y := x - \epsilon z) \\ &= \int_{\mathbb{R}^n} M(z) f(x - \epsilon z)\, dz. \end{aligned}$$

Since $M \in L^1(\mathbb{R}^n)$ and in view of (43) this implies that $\lim_{\epsilon\to 0} M_\epsilon * f(x) = f(x)$ for almost every $x \in \mathbb{R}^n$.

To conclude, we derive (44) :

$$\begin{aligned} \mathcal{F}^*(m_\epsilon(\xi)(\mathcal{F}f)(\xi)) &= \int_{\mathbb{R}^n} e^{2\pi i x\xi} m(\epsilon\xi)\mathcal{F}f(\xi)\, d\xi \\ &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{2\pi i x\xi} e^{-2\pi i x'\xi} m(\epsilon\xi) f(x')\, dx' d\xi \qquad \text{(Fubini)} \\ &= \int_{\mathbb{R}^n} \left\{ \int_{\mathbb{R}^n} e^{2\pi i (x-x')\xi} m(\epsilon\xi) d\xi \right\} f(x')\, dx' \\ &= \int_{\mathbb{R}^n} M_\epsilon(x - x') f(x')\, dx' \end{aligned}$$

where

$$\begin{aligned} M_\epsilon(x) &= \int_{\mathbb{R}^n} e^{2\pi i x\xi} m(\epsilon\xi) d\xi \qquad\qquad (\xi = \eta/\epsilon) \\ &= \epsilon^{-n} \int_{\mathbb{R}^n} e^{2\pi i \frac{x}{\epsilon}\eta} m(\eta)\, d\eta \\ &= \epsilon^{-n} \mathcal{F}^* m(x/\epsilon). \end{aligned}$$

□

75. Tempered distributions

We have defined the Fourier transform for L^1 functions. Can one define $\mathcal{F}f$ if f is merely a distribution? Following the Distribution Way of Doing Things we can try to define $\mathcal{F}T$ for any distribution by trying to make sense of $\langle \mathcal{F}T, \varphi\rangle$ for any test function φ.

Lemma 92. *For $f, g \in L^1(\mathbb{R}^n)$ one has*

$$\langle \mathcal{F}f, g\rangle = \langle f, \mathcal{F}g\rangle \tag{45}$$

where we write $\langle f, g\rangle = \int_{\mathbb{R}^n} f(x)g(x)\, dx$ whenever $f \cdot g \in L^1(\mathbb{R}^n)$.

Proof. Both sides equal

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{-2\pi i x\xi} f(\xi) g(x)\, d\xi\, dx.$$

□

We would therefore like to define $\langle \mathcal{F}T, \varphi\rangle = \langle T, \mathcal{F}\varphi\rangle$. But we can't do this for an arbitrary $T \in \mathcal{D}'(\mathbb{R}^n)$ since $\mathcal{F}\varphi$ is in general not a compactly supported smooth function, i.e. for most testfunctions φ the Fourier transform $\mathcal{F}\varphi$ is not a testfunction. A remedy for this problem is to enlarge the class of test functions. This leads us to the definition of *rapidly decreasing functions* (or Schwarz functions, after Laurent Schwartz who introduced them).

Definition. A function $f : \mathbb{R}^n \to \mathbb{C}$ is called *rapidly decreasing* if for all $k \in \mathbb{N}$ and $|\alpha| \leq k$ one has

$$p_{k,\alpha}(f) = \sup_{x \in \mathbb{R}^n} \left(1 + |x|\right)^k |D^\alpha f(x)| < \infty.$$

Thus rapidly decreasing functions are functions whose derivatives of arbitrary order decay faster than $(1+|x|)^{-N}$ as $|x| \to \infty$, for all N. The space of rapidly decreasing functions is denoted by $\mathcal{S}(\mathbb{R}^n)$, or just $\mathcal{S}$ if the dimension is clear from context.

Lemma 93. *For any $f \in \mathcal{S}$ one has $\mathcal{F}f, \mathcal{F}^*f \in \mathcal{S}$.*
For any $f \in \mathcal{S}$ one has $\mathcal{F}^*\mathcal{F}f = \mathcal{F}\mathcal{F}^*f = f$ pointwise.

Proof. Let $f \in \mathcal{S}$ be given. Since f and all its derivatives decrease faster than any $(1 + |x|)^{-N}$ as $|x| \to \infty$, we may use the basic properties of the Fourier transform to conclude

$$\xi^\beta D^\alpha \mathcal{F}f(\xi) = (-1)^{|\alpha|}(-2\pi i)^{|\alpha|-|\beta|}\mathcal{F}(D^\beta x^\alpha f).$$

Thus $\xi^\beta D_\xi^\alpha \mathcal{F}f(\xi)$ is bounded for arbitrary α, β. It follows that $\mathcal{F}f \in \mathcal{S}$.

The same arguments apply to $\mathcal{F}^*f$.

Since $\mathcal{S} \subset L^1$ we have proved in particular that for any $f \in \mathcal{S}$ one has $f, \mathcal{F}f, \mathcal{F}^*f \in L^1(\mathbb{R}^n)$ so that our inversion Theorem 91 applies. □

Definition. A tempered distribution is a linear functional $T : \mathcal{S} \to \mathbb{C}$ which is continuous in the sense that for some K one has

$$|T(\varphi)| \leq C \sup_{\substack{x \in \mathbb{R}^n \\ |\alpha| \leq K}} \left(1 + |x|\right)^K |D^\alpha \varphi(x)|.$$

The space of tempered distributions is denoted by $\mathcal{S}'$.

The Fourier transform of a tempered distribution $T \in \mathcal{S}'$ is given by

$$\langle \mathcal{F}T, \varphi \rangle \stackrel{\text{def}}{=} \langle T, \mathcal{F}\varphi \rangle .$$

Similarly, one defines

$$\langle \mathcal{F}^*T, \varphi \rangle \stackrel{\text{def}}{=} \langle T, \mathcal{F}^*\varphi \rangle .$$

Theorem 94 (Inversion for tempered distributions.)**.** *For any $T \in \mathcal{S}'$ one has $\mathcal{F}^*\mathcal{F}T = T$.*

Proof. We have observed that for all Schwarz functions $\varphi \in \mathcal{S}$ one has $\mathcal{F}\mathcal{F}^*\varphi = \varphi$. Hence one has

$$\langle \mathcal{F}^*\mathcal{F}T, \varphi \rangle = \langle \mathcal{F}T, \mathcal{F}^*\varphi \rangle = \langle T, \mathcal{F}\mathcal{F}^*\varphi \rangle = \langle T, \varphi \rangle .$$

□

Exercise 139. Which of the following functions f define tempered distributions T_f, if one sets

$$\langle T_f, \varphi \rangle = \lim_{A,B \to \infty} \int_{-B}^{A} f(x)\varphi(x)dx?$$

(i) $f(x) = |x|$;
(ii) $f(x) = e^{|x|}$;
(iii) $f(x) = e^x \sin e^x$.

Exercise 140. One can define differentiation of tempered distributions in the same way as for ordinary distributions, namely,

$$\langle D_j T, \varphi \rangle \stackrel{\text{def}}{=} -\left\langle T, \frac{\partial \varphi}{\partial x_j} \right\rangle$$

for any $\varphi \in \mathcal{S}, T \in \mathcal{S}'$. This gives us two possibly conflicting interpretations of "$f = D_j g$ in the sense of distributions" when f and g are locally integrable functions.

Suppose $f, g \in L^p(\mathbb{R}^n)$ for some p. *Show that $f = D_j g$ in $\mathcal{D}'$ if and only if $f = D_j g$ in $\mathcal{S}'$.*

76. Plancherel's Formula

Lemma 95. *For $f, g \in \mathcal{S}$ one has $(\mathcal{F}f, g) = (f, \mathcal{F}^* g)$, where (f, g) is the $L^2(\mathbb{R}^n)$ inner product.*

Proof. Remembering the complex conjugate in the definition of the inner product on L^2 one finds $(f, g) = \langle f, \bar{g} \rangle$. Hence

$$(\mathcal{F}f, g) = \langle \mathcal{F}f, \bar{g} \rangle = \langle f, \mathcal{F}\bar{g} \rangle = \langle f, \overline{\mathcal{F}^* g} \rangle = (f, \mathcal{F}^* g).$$

□

By setting $g = \mathcal{F}f$ one finds *Plancherel's formula*

$$(\mathcal{F}f, \mathcal{F}f) = (f, \mathcal{F}^* \mathcal{F} f) = (f, f),$$

in other words, the Fourier transform preserves the L^2 norm on rapidly decaying functions! Since $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is bijective (its inverse is given by $\mathcal{F}^*$) the same is true for $\mathcal{F}^* : \mathcal{S} \to \mathcal{S}$. Since $\mathcal{S}$ is dense in L^2 we conclude

Theorem 96. *The Fourier transform $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ extends to an norm preserving linear map $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. This map is invertible, its inverse is given by $\mathcal{F}^*$.*

For any $f \in L^2(\mathbb{R}^n)$ one has $\mathcal{F}^*\mathcal{F}f = f$, but one should be careful with the interpretation of this identity. Here both sides do not stand for continuous functions which are well defined at each $x \in \mathbb{R}^n$. Since $L^2(\mathbb{R}^n) \not\subset L^1(\mathbb{R}^n)$ we have no guarantee that the integral

$$\mathcal{F}f(x) = \int_{\mathbb{R}^n} e^{-2\pi i x\xi} f(x)\, dx$$

exists for any $\xi \in \mathbb{R}^n$. Instead, we have defined $\mathcal{F}f$ by

$$\mathcal{F}f = (L^2) \lim_{k \to \infty} \mathcal{F} f_k$$

where $f_k \in \mathcal{S}$ is a sequence of Schwarz functions which converge in L^2 to f. The result is an L^2 function $g = \mathcal{F}f$. Indeed, Theorem 96 says that the result can be *any* function $g \in L^2$!

77. Fourier multiplier operators.

A Fourier multiplier operator T_m is an operator of the form

$$T_m f = \mathcal{F}^*\big(m(\xi)\mathcal{F}^* f(\xi)\big)$$

for some bounded function $m : \mathbb{R}^n \to \mathbb{C}$ which is called the multiplier.

If m happens to be the Fourier transform of some $M \in L^1(\mathbb{R}^n)$ then

$$T_m f = \mathcal{F}^*\big(m(\xi)\mathcal{F}f(\xi)\big) = \mathcal{F}^*\big(\mathcal{F}M(\xi)\mathcal{F}f(\xi)\big) = \mathcal{F}^*\mathcal{F}(M * f) = M * f.$$

In other words for $m = \mathcal{F}M(\xi)$ the Fourier multiplier operator T_m is nothing but convolution with an L^1 function M. Such operators are bounded on all $L^p(\mathbb{R}^n)$ with $1 \leq p \leq \infty$, by Young's inequality.

Theorem 97. *If $m \in L^\infty(\mathbb{R}^n)$ then the operator T_m is bounded on $L^2(\mathbb{R}^n)$.*

Proof. If $f \in L^2$ then $\mathcal{F}f \in L^2$, with $\|\mathcal{F}f\|_{L^2} = \|f\|_{L^2}$. Since $m \in L^\infty$ one has $m(\xi)\mathcal{F}f(\xi) \in L^2$ with

$$\|m(\xi)\mathcal{F}f(\xi)\|_{L^2} \leq \|m\|_{L^\infty}\|f\|_{L^2}.$$

Since the inverse transform also preseves the L^2 norm one obtains

$$T_m f = \mathcal{F}^*\big(m(\xi)\mathcal{F}f(\xi)\big) \in L^2$$

with

$$\|T_m f\|_{L^2} = \big\|\mathcal{F}^*\big(m(\xi)\mathcal{F}f(\xi)\big)\big\|_{L^2} = \|m(\xi)\mathcal{F}f(\xi)\|_{L^2} \leq \|m\|_{L^\infty}\|f\|_{L^2}.$$

□

This is a very strong theorem since nothing is required of the multiplier besides being measurable and (essentially) bounded. However, the theorem is only true in this generality for operators from L^2 to L^2. Similar theorems exist giving sufficient conditions on the multiplier m for the operator T_m to be bounded on L^p. On the other hand, a simple example exists of a multiplier whose operator T_m is not bounded on any L^p except $p = 2$, namely: let m be the characteristic function of the unit ball in $\mathbb{R}^n$ (the example is simple, the proof is not – see [11, page 450].)

78. An example of Elliptic Regularity

The Fourier transform can be used to prove that solutions of certain partial differential equations have more derivatives than is apparent from the equation. In this section we illustrate this with the Laplace operator. As a result we will show that the eigenfunctions of the Laplacian on a domain $\Omega \subset \mathbb{R}^n$ are in fact C^∞ functions in the domain.

78.1. The Resolvent of the Laplacian.

For given $\lambda \in \mathbb{C}$ we try to solve the equation

$$\lambda u - \Delta u = f \tag{46}$$

where the given function f and the solution u shall be tempered distributions.

Lemma 98. *If $\lambda \in \mathbb{C} \setminus (-\infty, 0]$ then (46) has a unique solution $u \in L^2(\mathbb{R}^n)$ for any $f \in L^2(\mathbb{R}^n)$. This solution is given by*

$$u = \mathcal{F}^*\big\{m_\lambda \cdot \mathcal{F}f\big\},$$

where the multiplier m_λ is given by

$$m_\lambda(\xi) = \frac{1}{1 + 4\pi^2|\xi|^2}.$$

Furthermore, the first and second derivatives of u in the sense of distributions also belong to L^2.

Proof. Taking Fourier transforms on both sides one finds

$$\mathcal{F}f(\xi) = \mathcal{F}(\lambda u - \Delta u) = \lambda \mathcal{F}u(\xi) + 4\pi^2|\xi|^2 \mathcal{F}u(\xi),$$

which one can solve easily for $\mathcal{F}u(\xi)$ by dividing by $\lambda + 4\pi^2|\xi|^2$, i.e. by multiplying with m_λ.

If $\lambda \in \mathbb{C} \setminus (-\infty, 0]$ then m_λ is a smooth and bounded function on $\mathbb{R}^n$, so that the corresponding multiplier operator T_{m_λ} is bounded on L^2.

To see that the derivatives of u are L^2 functions we observe that

$$\mathcal{F}(D_k D_l u) = -4\pi^2 \xi_k \xi_l \mathcal{F}u(\xi) = \frac{-4\pi^2 \xi_k \xi_l \mathcal{F}u(\xi)}{\lambda + 4\pi^2|\xi|^2} \mathcal{F}f(\xi).$$

Thus if $f \in L^2$ then by Parseval we have $\mathcal{F}f \in L^2$; the multiplier above is bounded, so $\mathcal{F}(D_k D_l u)$ also belongs to L^2; by Parseval again we see that $D_k D_l u$ itself also belongs to L^2.

A similar argument shows that the first derivatives also belong to L^2. □

Recall that $W^{m,p}(\mathbb{R}^n)$ is the Sobolev space of functions whose distributional derivatives of order $\leq m$ are L^p functions.

Corollary 99. *If $u \in L^2$, and if $\Delta u \in W^{m,2}(\mathbb{R}^n)$ then $u \in W^{m+2,2}(\mathbb{R}^n)$.*

In other words, "if the Laplacian of an L^2 function has m derivatives in L^2 then the function has $m+2$ derivatives in L^2."

Proof. By induction on $m \geq 0$.

For $m = 0$ we are given $u \in L^2$ and $\Delta u \in L^2$. Hence $u - \Delta u \in L^2$. By lemma 98 the first and second derivatives of u belong to L^2. This means $u \in W^{2,2}(\mathbb{R}^n)$.

In the induction step with $m > 0$ we are given $u \in L^2$ with $\Delta u \in W^{m,2}$. In particular $\Delta u \in L^2$ so by the $m = 0$ step we conclude that $u \in W^{2,2}$, i.e. the derivatives of order ≤ 2 are in L^2.

Consider $v_k = D_k u$. Then we have just shown that $v_k \in L^2$. Since $\Delta v_k = D_k \Delta u$, we also have $\Delta v_k \in W^{m-1,2}$ (i.e. v_k has derivatives of order $\leq m-1$ in L^2). The induction hypothesis then implies that $v_k \in W^{m+1,2}$ (i.e. v_k has derivatives of order $\leq m+1$ in L^2). Recall that the v_k are the first derivatives of u, so that we have $u \in W^{m+2,2}$, as claimed. □

78.2. Smoothness of Eigenfunctions

In this section we study local smoothness of the eigenfunctions of the Laplace operator. We will use the following definitions:

Let $\Omega \subset \mathbb{R}^n$ be an open domain. A distribution $u \in \mathcal{D}'(\Omega)$ is said to be locally of class $W^{m,2}$ if for every $\phi \in \mathcal{D}(\Omega)$ one has $\phi u \in W^{m,2}(\mathbb{R}^n)$. We write $W^{m,2}_{\text{loc}}(\Omega)$ for the space of locally $W^{m,2}$ distributions on Ω.

Theorem 100. *Let $\Omega \subset \mathbb{R}^n$ be an open domain, and let $u \in W^{1,2}_{\text{loc}}(\Omega)$ be a solution of $\Delta u = \lambda u$ for some $\lambda \in \mathbb{C}$. Then $u \in C^\infty(\Omega)$.*

We will break the proof into several Lemmas.

Lemma 101. *Let $u \in W^{1,2}_{\text{loc}}(\Omega)$ be a solution of $\Delta u = \lambda u$ in the sense of distributions. Then $u \in W^{m,2}_{\text{loc}}(\Omega)$ for all m, i.e. for any test function $\phi \in \mathcal{D}(\Omega)$ one has $\phi u \in W^{m,2}(\mathbb{R}^n)$.*

Proof. By induction on m. The case $m = 1$ is given, so we proceed directly to the induction step. Assume we have proven that $u \in W^{m,2}_{\text{loc}}(\Omega)$.

Let $\phi \in \mathcal{D}(\Omega)$ be given. Then

$$\Delta(\phi u) = (\Delta u)\phi + 2\nabla u \cdot \nabla\phi + u\Delta\phi = 2\nabla u \cdot \nabla\phi + u(\Delta\phi + \lambda\phi).$$

By the induction hypothesis both $\nabla u \cdot \nabla\phi$ and $u(\Delta\phi + \lambda\phi)$ belong to $W^{m-1,2}(\mathbb{R}^n)$. Hence $\phi u, \Delta(\phi u) \in W^{m,2}(\mathbb{R}^n)$ and we conclude that $\phi u \in W^{m+1,2}(\mathbb{R}^n)$. □

Lemma 102. *If $m > \frac{n}{2} + k$ and $u \in W^{m,2}(\mathbb{R}^n)$, then $u \in C^k(\mathbb{R}^n)$ and the derivatives of order k of u are Hölder continuous with exponent*

Proof. One can derive this from the Sobolev embedding theorems. Here is a proof which uses the Fourier transform.

We are given $u \in W^{m,2}(\mathbb{R}^n)$, i.e. all derivatives $D^\alpha u$ of order $|\alpha| \le m$ belong to L^2. Their Fourier transforms $\mathcal{F}(D^\alpha u) = (-2\pi\xi)^\alpha \mathcal{F}u$ therefore also lie in $L^2(\mathbb{R}^n)$. We therefore have

$$\int_{\mathbb{R}^n} (1 + |\xi|^2)^m |\mathcal{F}u(\xi)|^2 \, d\xi < \infty. \tag{47}$$

The inversion theorem tells us that

$$u(x) = \int_{\mathbb{R}^n} e^{2\pi i x\xi} \mathcal{F}u(\xi) \, d\xi$$

Condition (47) implies via Hölder's inequality that

$$\begin{aligned}\int_{\mathbb{R}^n} |\xi|^k |\mathcal{F}u(\xi)| \, d\xi &= \int_{\mathbb{R}^n} (1 + |\xi|^2)^{-m/2} |\xi|^k (1 + |\xi|^2)^{m/2} |\mathcal{F}u(\xi)| \, d\xi \\ &\le \left(\int_{\mathbb{R}^n} (1 + |\xi|^2)^{-m} |\xi|^{2k} \, d\xi \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} (1 + |\xi|^2)^m |\mathcal{F}u(\xi)|^2 \, d\xi \right)^{\frac{1}{2}}\end{aligned}$$

The condition $k > m + \frac{n}{2}$ implies that the first integral here converges. We may therefore differentiate k times under the integral to obtain for any α with $|\alpha| = k$ that

$$D^\alpha_x u(x) = \int_{\mathbb{R}^n} e^{2\pi i x\xi} (2\pi i\xi)^\alpha \mathcal{F}u(\xi) \, d\xi$$

so $D^\alpha u$ is the inverse Fourier transform of an integrable function. Hence $D^\alpha u$ is continuous. □

References

[1] Yu.D.Burago & V.A.Zalgaller, Geometric Inequalities, Springer Grundlehren **285** (1988).

[2] L.C.Evans, Partial Differential Equations, A.M.S. Graduate Studies in Mathematics, **19** (1998).

[3] F.G.Friedlander, Introduction to the theory of distributions, Cambridge University Press, 1982.

[4] W.Rudin, Real and Complex Analysis.

[5] W.Rudin, Functional Analyis.

[6] G.Bachman and L.Narici, Functional Analysis, Academic Press 1966.

[7] E.M.Stein, Singular integrals and Differentiability Properties of Functions, Princeton University Press (1970)

[8] R.J.Zimmer, Essential results of Functional Analysis, Chicago Lectures in Mathematics, University of Chicago Press (1990).

[9] P.Halmos, V.Sunder, Bounded INtegral operators on L^2 spaces, Springer Verlag.

[10] Y.Katznelson, An Introduction to Harmonic Analysis, Dover Publications, 1976.

[11] E.M.Stein, Harmonic Analysis – Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press (1993)