

Benjamin McKay

Introduction to Partial Differential Equations

Sobolev Spaces for Linear Elliptic Equations

May 1, 2020

Preface

These lecture notes provide proofs of some elementary results about linear partial differential equations in domains in Euclidean space. Chapter 1 is a review of the prerequisite material from several variable calculus, and also provides the definitions, and statements of theorems (without proof) of the main results in measure theory and integration. The main theorem is the dominated convergence theorem: pointwise convergence of functions, all smaller than an integrable function, ensures convergence of integrals. Limits of Riemann sums are poorly behaved under pointwise limits of functions, so the dominated convergence theorem requires the more sophisticated Lebesgue integration.

Chapters 2 to 3 survey some elementary results about approximation of rough functions by smooth functions. We often need to allow solutions of partial differential equations to be poorly behaved functions, for example in modelling shock waves or explosions. But differential equations are expressed in terms of derivatives, which only exist for relatively smooth functions. A large part of our effort is aimed at resolving this paradox. Some functions are not differentiable strictly speaking, but still behave very much as if they had derivatives. Expressing a rough function as a limit of smooth functions, we can think of its derivative as a limit of derivatives of smooth functions. If this limit exists in a suitable sense, it is called a *weak derivative*. It can be easier to solve partial differential equations using weak derivatives. A function is Sobolev of order k if the function and its various weak derivatives up to order k have well enough behaved integrals. We will mostly search for solutions to partial differential equations among Sobolev functions.

When searching for solutions to partial differential equations, we might hope to explicitly write them down with some formulas. This is rarely possible, but the cases where we succeed are vital sources of intuition. When this fails, we might instead construct a scheme which starts with a guess, an approximate solution, which we can write down, and replace it with a better guess, repeatedly, aiming to converge to a solution. The Kondrashov–Rellich compactness theorem tells us when a sequence of Sobolev functions converges to a Sobolev function. We can often use this to prove convergence of our scheme. We then need to bridge the gap between derivatives in the weak sense and derivatives in the usual sense of calculus. The Sobolev embedding theorem states that all Sobolev functions of high enough order have some number of derivatives which are not just weak derivatives. If we can find solutions of partial differential equations in the weak sense (for instance by the Kondrashov–Rellich theorem), then Sobolev’s embedding theorem might tell us that they are actually solutions in

the usual sense.

Contents

1	Analysis Review	1
2	Approximation and Convolution	11
3	Sobolev spaces	19
4	Fourier Transforms	29
5	Distributions	37
6	L^2 Theory of Derivatives	41
7	The Direct Method of the Calculus of Variations	47
8	Linear Elliptic Second Order Partial Differential Equations	55
9	Pseudodifferential Operators	61
	Bibliography	73
	List of notation	75
	Index	77

Chapter 1

■ Analysis Review

We go over terminology and notation from analysis, including a few results which you might not have already covered.

Euclidean space

We use the usual terminology and notation of sets without introduction. We write \mathbb{R} to mean the set of all real numbers, \mathbb{C} the set of complex numbers. Suppose that $f: X \rightarrow Y$ is a map between sets and $S \subset X$ is a subset. The *image* $f(S)$ of S is the set of all points $f(x)$ for all $x \in S$. The *image* of f is $f(X)$. Similarly if $T \subset Y$ is a subset, the *preimage*, $f^{-1}T$, of T is the set of points $x \in X$ for which $f(x) \in T$. It will often be convenient to avoid choosing a name for a function, for example writing $x \mapsto x^2 \sin x$ to mean the function $f: \mathbb{R} \rightarrow \mathbb{R}, f(x) = x^2 \sin x$.

The set \mathbb{R}^n is the set of all n -tuples

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

of real numbers $x_1, x_2, \dots, x_n \in \mathbb{R}$. Following standard practice, we will often be lazy and write such a tuple horizontally as

$$x = (x_1, x_2, \dots, x_n).$$

We refer to \mathbb{R}^n as *Euclidean space* and to its elements as either *points* (in which case we draw them as dots) or as *vectors* (in which case we draw them as arrows from the origin). Similarly \mathbb{C}^n is the set of all n -tuples of complex numbers. We refer to \mathbb{C}^n as *complex Euclidean space*.

If $x, y \in \mathbb{R}^n$, their *inner product* or *scalar product* or *dot product* is

$$\langle x, y \rangle = \sum_i x_i y_i.$$

If $z, w \in \mathbb{C}^n$, their inner product is

$$\langle z, w \rangle = \sum_a z_a \bar{w}_a.$$

The *length* of a vector $x \in \mathbb{R}^n$ or $z \in \mathbb{C}^n$ is

$$\|x\| = \sqrt{\langle x, x \rangle}, \quad \|z\| = \sqrt{\langle z, z \rangle}.$$

The *distance* between two points $x, y \in \mathbb{R}^n$ is $d(x, y) = \|x - y\|$. The *ball* or *open ball* of radius r around a point $x \in \mathbb{R}^n$ is the set $B_r(x)$ of all points of \mathbb{R}^n of distance less than r from x . The *closed ball* of radius r around a point $x \in \mathbb{R}^n$ is the set $\bar{B}_r(x)$ of all points $p \in \mathbb{R}^n$ of distance less than or equal to r from x . A set is *bounded* if it lies in a ball. A map is *bounded* if its image is bounded.

1.1 Prove that every box is bounded.

A set $U \subset \mathbb{R}^n$ is *open* if it is a union of open balls. The *closure* \bar{S} of a set $S \subset \mathbb{R}^n$ is the set of all points p so that any open set containing p contains points of S . The *boundary* is the set of points p so that any open set around p contains points of S and points outside of S , i.e. $\partial S = \bar{S} \cap \overline{\mathbb{R}^n - S}$. A *domain* is an open set $D \subset \mathbb{R}^n$ so that $\partial D = \partial(\mathbb{R}^n - \bar{D})$.

1.2 Prove that every open ball is a domain.

1.3 Give an example of an open set which is *not* a domain.

For any real numbers $a, b \in \mathbb{R}$, we write $[a, b] \subset \mathbb{R}$ to mean the set of points $x \in \mathbb{R}$ so that $a \leq x \leq b$, and we call $[a, b]$ the *closed interval* from a to b , etc. A *box* in \mathbb{R}^n is a subset of the form

$$X = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n],$$

i.e. a product of closed intervals. In other words, a vector $x \in \mathbb{R}^n$ lies in the box X just when $a_1 \leq x_1 \leq b_1$ and $a_2 \leq x_2 \leq b_2$ and ... and $a_n \leq x_n \leq b_n$.

A set $S \subset \mathbb{R}^n$ is *compact* if it is closed and bounded. The image of a compact set under a continuous map is compact. A *cover* of a set S is a collection of sets, say X_a for $a \in A$, so that every point of S lies in at least one of these sets X_a . For any cover U_a of a compact set S by open sets, S is already covered by a finite collection of those open sets.

Derivatives

Write ∂_i to mean $\frac{\partial}{\partial x_i}$, and similarly write ∂_x to mean $\frac{\partial}{\partial x}$, and ∂_{ij} to mean $\frac{\partial^2}{\partial x_i \partial x_j}$ and so on. Write df for the “differential”

$$df = (\partial_1 f, \partial_2 f, \dots, \partial_n f),$$

which we can clearly write as

$$df = \sum_i \frac{\partial f}{\partial x_i} dx_i.$$

If $a = (a_1, a_2, \dots, a_n)$, let

$$\partial^a = \frac{\partial^{a_1}}{\partial^{a_1} x_1} \frac{\partial^{a_2}}{\partial^{a_2} x_2} \cdots \frac{\partial^{a_n}}{\partial^{a_n} x_n}.$$

In particular, $\partial^0 f = f$. A function $f: U \rightarrow \mathbb{R}^p$ defined on an open set $U \subset \mathbb{R}^n$ is C^k if all derivatives $\partial^a f$ are defined and continuous for all $|a| \leq k$. A function is C^∞ , also called *smooth*, if it is C^k for all k . A function f defined on *any* set is C^k if, near each point where f is defined, f can be somehow (in many ways) extended to a C^k function in some open set around that point (maybe in different ways around different points). For functions valued in real or complex numbers or vectors, write $f(x) = o(g(x))$ to mean that $\frac{\|f(x)\|}{\|g(x)\|} \rightarrow 0$ as $x \rightarrow 0$. Similarly, write $f(x) = o(g(x))^k$ to mean that $\frac{\|f(x)\|}{\|g(x)\|^k} \rightarrow 0$ as $x \rightarrow 0$. If $a = (a_1, a_2, \dots, a_n)$ with each $a_j \geq 0$ an integer, let $|a| = a_1 + a_2 + \cdots + a_n$, $a! = a_1! a_2! \cdots a_n!$ and for $x \in \mathbb{R}^n$ let $x^a = x_1^{a_1} x_2^{a_2} \cdots x_n^{a_n}$. Every C^k function has a Taylor series

$$f(x) = \sum_a \partial^a f(0) \frac{x^a}{a!} + o(x)^k.$$

A function f is $C^{0,\alpha}$ if for every compact set K on which f is defined, there is a constant $C > 0$ so that if x and y lie in K and x and y are close enough, then $d(f(x), f(y)) \leq C d(x, y)^\alpha$. Say that f is $C^{k,\alpha}$ if $\partial^a f$ is $C^{0,\alpha}$ for all $|a| \leq k$. Similarly, if f is not defined on an open set, we can say f is $C^{k,\alpha}$ if, near each point where f is defined, f can be extended to a $C^{k,\alpha}$ function in some open set containing that point. For any set $X \subset \mathbb{R}^n$, let $C^k(X)$ be the set of all C^k functions on X , and let $C^{k,\alpha}(X)$ be the set of all $C^{k,\alpha}$ functions on X . A function f is C_b^k if it is C^k with all derivatives of order up to k bounded. The norm of such a function is

$$\|f\|_{C^k} = \sup_{|a| \leq k} |\partial^a f|.$$

In this norm, if $X = \bar{U}$ is the closure of a bounded open set, then both of $C^k(X) \subset C_b^k(U)$ are complete metric spaces. For any $f \in C^{k,\alpha}$, let

$$\|f\|_{C^{k,\alpha}} = \sum_{|a| \leq k} \sup_{x \neq y} \frac{d(\partial^a f(x), \partial^a f(y))}{d(x, y)^\alpha}.$$

In this norm, if $X = \bar{U}$ is the closure of a bounded open set, then $C^{k,\alpha}(X)$ is a complete metric space.

1.4 If $\alpha > 1$ prove that $f \in C^{0,\alpha}$ just when f is constant.

A sequence of functions f_1, f_2, \dots is *equicontinuous* if, for any point s , for every $\varepsilon > 0$, there is a neighborhood of s so that for every t in that neighborhood, all of the differences $f_1(s) - f_1(t), f_2(s) - f_2(t), \dots$ are smaller in absolute value than ε .

1.5 Give an example of a sequence of continuous functions which is not equicontinuous.

Theorem 1.1 (Ascoli–Arzelà). *If f_1, f_2, \dots is an equicontinuous sequence of functions and $|f_1|, |f_2|, \dots$ are all bounded by the same constant then some subsequence converges uniformly.*

1.6 Suppose that f_1, f_2, \dots is a sequence of $C^{0,\alpha}$ functions of bounded norm. Prove that for every $\beta < \alpha$, all of the functions f_1, f_2, \dots belong to $C^{0,\beta}$, and that there is a convergent subsequence in $C^{0,\beta}$.

Smooth functions with compact support

The *support* of a function is the closure of the set of points where the function doesn't vanish. A smooth function is called a *test function* if it has compact support. Write C_c^∞ for the set of test functions. Note that if D is a domain, a function on D with compact support must vanish near every point of ∂D . We say that a sequence $f_1, f_2, \dots \in C_c^\infty(U)$ converges to an element $f \in C_c^\infty(U)$ if there is a compact set $K \subset U$ containing the supports of all elements of the sequence and $\partial^a f_1, \partial^a f_2, \dots \rightarrow \partial^a f$ uniformly, for every multiindex a . For any two concentric spheres, there is a test function on \mathbb{R}^n equal to 1 inside the smaller sphere, and equal to 0 outside the larger sphere, symmetric under rotations around the common centre of the spheres, and strictly decreasing along every radial line out of the centre. One can write down an explicit example of such a function, via a long but straightforward exercise in cutting and pasting, using the fact that the function $f(x) = e^{-1/x}$ vanishes at the origin with all derivatives.

An *open cover* of a set $S \subset \mathbb{R}^n$ is a collection of open sets $U_\alpha \subset \mathbb{R}^n$ so that $S \subset \bigcup_\alpha U_\alpha$. A *partition of unity* on a set $S \subset \mathbb{R}^n$ is a collection of smooth functions $f_\alpha: \mathbb{R}^n \rightarrow [0, 1]$ so that

- a. each point x lies in an open set on which only finitely many of the functions f_α are not everywhere zero and
- b. $\sum_\alpha f_\alpha(x) = 1$ at every point $x \in S$.

The partition of unity is *subordinate* to an open cover $\{U_\alpha\}$ if every f_α is supported in a compact subset of U_α . If S is a closed set, then every open cover of S has a partition of unity subordinate to it.

Measure

We will need some facts about Lebesgue measure which you might not have run into yet; we will just summarise the relevant facts and not prove any of them. See [1, 2] for excellent introductions. Paradoxically, there is no reasonable way to assign a volume to every subset of Euclidean space, or to associate an integral to every function, so we assign volumes and integrals to various sets and to

various functions; luckily among those sets and functions one finds any set or function that we can explicitly describe or would ever need to think about. The *length* of an interval $[a, b]$ or (a, b) or $[a, b)$ or $(a, b]$ is $b - a$. If we write a box as a product of intervals, the *volume* of the box is the product of the lengths of the intervals. Picture a set S covered by a collection of boxes. By perhaps replacing these boxes by some smaller ones, we can try to cover S without very much overlap. The *outer measure* of a set $S \subset \mathbb{R}^n$ is the number V so that we can cover S by a sequence of boxes whose sum of volumes can be as close as we like to V , but can't be less than V . The outer measure of a box turns out to equal its volume. The graph of a map, say $f: X \rightarrow Y$, where $X \subset \mathbb{R}^p$ and $Y \subset \mathbb{R}^q$, is clearly a subset of \mathbb{R}^{p+q} . If a set $S \subset \mathbb{R}^n$ lies in the graph of a continuous map then S has outer measure zero. A set $S \subset \mathbb{R}^n$ is *measureable* if it can be approximated well by open sets, in the sense that there are open sets U containing S so that $U - S$ has outer measure as small as we like; clearly open sets are measureable. The complement of any measureable set is measureable. The union and the intersection of any sequence of measureable sets is measureable. The outer measure of a measureable set is called its *measure*. If a set has outer measure zero, then it is measureable (with measure zero). Given a sequence of disjoint measureable sets, the measure of the union is the sum of the measures. If we say that a statement about a point x is true “almost everywhere”, we mean that the counterexamples form a set of measure zero.

Integration

We will henceforth treat any two functions as being the same if they agree everywhere except on a set of measure zero. This has the strange consequence that when we say a function is continuous or differentiable or bounded, we really mean that it can be made continuous or differentiable or bounded, after we modify it on a set of measure zero. When we refer to an upper bound on a function f , written $\sup_x f(x)$, we really mean the smallest value y_0 so that the set of points x on which $f(x) > y_0$ has measure zero. We will also be deliberately vague about whether our functions are real-valued or complex-valued. A function f is *Riemann integrable* if the usual limits of lower and upper Riemann sums agree. A function f is *measureable* if, for any number a , the set of points x at which $f(x) < a$ is measureable. Suppose that f is a measureable and nonnegative function on a measureable set $X \subset \mathbb{R}^n$. For each t , let $f^*(t)$ be the measure of $f^{-1}(t, \infty)$. Then f^* is a decreasing function on the real number line, and so is Riemann integrable; define $\int f = \int_{-\infty}^{\infty} f^*(t) dt$. A function f is *integrable* if it is measureable and $|f|$ has finite integral. If f is real-valued, it follows that the functions $f_+(x) = \max(0, f(x))$ and $f_-(x) = \max(0, -f(x))$ are integrable, and we let $\int f = \int f_+ - \int f_-$. Similarly, for f complex-valued $f = g + ih$ with g and h real-valued, set $\int f = \int g + i \int h$, and similarly if f is vector-valued.

Theorem 1.2 (The dominated convergence theorem). *If f_1, f_2, \dots is a sequence of integrable functions on some measureable set, converging pointwise, and $|f_1|, |f_2|, \dots$ are all bounded by the same integrable function, then the pointwise limit $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ is integrable and*

$$\int f_n \rightarrow \int f \text{ as } n \rightarrow \infty.$$

Theorem 1.3 (Fubini). *Suppose that $X \subset \mathbb{R}^p$ and $Y \subset \mathbb{R}^q$ are measureable sets. If f is integrable on $X \times Y$ then $y \mapsto f(x, y)$ is integrable for all $x \in X$ except on a measure zero set, and $x \mapsto f(x, y)$ is integrable for all $y \in Y$ except on a measure zero set and $\int_{X \times Y} f = \int_X \left(\int_Y f(x, y) dy \right) dx = \int_Y \left(\int_X f(x, y) dx \right) dy$.*

For any measureable set $X \subset \mathbb{R}^n$, let $L^1(X)$ be the set of all integrable functions on X . We often write L^1 if we wish to leave the particular choice of measureable set X unspecified; in that case we usually mean $X = \mathbb{R}^n$. For $p \geq 1$, let $L^p(X)$ be the set of measureable functions f for which $|f|^p \in L^1(X)$, and let $\|f\|_{L^p} = \left(\int_X |f|^p \right)^{1/p}$. Similarly, let $L^\infty(X)$ be the set of bounded measureable functions, and let $\|f\|_{L^\infty} = \sup |f|$ be the uniform norm. If $1 \leq p \leq \infty$, the distance between two functions $f, g \in L^p$ is $\|f - g\|_{L^p}$. With this notion of distance, the space L^p is a complete metric space, i.e. any Cauchy sequence converges, i.e. if $\|f_j - f_k\|_{L^p} \rightarrow 0$ as $j, k \rightarrow \infty$, then f_1, f_2, \dots converges. Moreover, the convergence of a sequence $u_j \rightarrow u$ in L^p implies that some subsequence converges pointwise almost everywhere, i.e. there is a subsequence u_{j_1}, u_{j_2} so that for almost every point x , $u_{j_k}(x) \rightarrow u(x)$. On any domain, the test functions are dense in L^p for $1 \leq p < \infty$, but *not* for $p = \infty$; the bounded C^∞ functions are dense in L^∞ .

Theorem 1.4 (Hölder's inequality). *The “inner product” $\langle f, g \rangle = \int f \bar{g}$ is defined for $f \in L^p$ and $g \in L^q$ as long as $\frac{1}{p} + \frac{1}{q} = 1$, or if $p = 1, q = \infty$ or $p = \infty, q = 1$, and*

$$|\langle f, g \rangle| \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Equality holds if and only if $a|f|^p = b|g|^q$ for some constants a, b , not both 0.

1.7 Suppose that X is a measureable set of finite measure. Prove that if $1 \leq p \leq q \leq \infty$ then $L^q(X) \subset L^p(X)$ and there is a number C so that for every function $f \in L^q(X)$, $\|f\|_{L^p} \leq C \|f\|_{L^q}$. On $X = [0, 1] \subset \mathbb{R}$ find the best (i.e. smallest possible) value for C .

1.8 Suppose that we have functions f_1, f_2, \dots, f_k on \mathbb{R}^n and that $f_j \in L^{p_j}$ and that

$$\frac{1}{p_1} + \frac{1}{p_2} + \dots + \frac{1}{p_k} = 1.$$

Use induction and Hölder's inequality to prove that

$$\int f_1 f_2 \dots f_k \leq \|f_1\|_{L^{p_1}} \|f_2\|_{L^{p_2}} \dots \|f_k\|_{L^{p_k}}.$$

Theorem 1.5 (Minkowski's inequality). *If $f, g \in L^p$, then $\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}$.*

1.9 Prove that for any $f \in L^p$, the linear map $g \in L^q \mapsto \langle g, f \rangle$ is a continuous linear map.

If $U \subset \mathbb{R}^n$ is open, write $L^p_{\text{loc}}(U)$ to mean the functions f so that the restriction of f to any bounded open set $W \subset U$ is in $L^p(W)$. The largest space of functions we will ever consider is L^1_{loc} , called the *locally integrable functions*. Convergence in L^p_{loc} means convergence of the restriction to U in $L^p(U)$ for every bounded open set U .

1.10 For each positive integer j , let d_j be the number of base 10 digits in j . Let f_1, f_2, \dots be the sequence of functions

$$f_j(x) = \begin{cases} 1, & \text{if } \frac{j}{10^{1+d_j}} \leq x \leq \frac{j+1}{10^{1+d_j}}, \\ 0, & \text{otherwise.} \end{cases}$$

Draw these functions. Which L^p spaces do these functions belong to? In which do they have a limit? In which is there a subsequence which has a limit?

Continuity of integrals and differentiation under the integral sign

The dominated converge theorem easily implies:

Theorem 1.6. *Suppose that $X \subset \mathbb{R}^p$ is a measurable set and $Y \subset \mathbb{R}^q$ is an arbitrary set, $f: X \times Y \rightarrow \mathbb{R}$, denoted $f(x, y)$, is integrable in x for each y and is continuous in y , and is bounded: $|f(x, y)| \leq |g(x)|$ for some integrable function g . Then $\int f(x, y) dx$ is continuous in y .*

Theorem 1.7. *Suppose that $X \subset \mathbb{R}^p$ is a measurable set and $Y \subset \mathbb{R}$ is an open interval, $f: X \times Y \rightarrow \mathbb{R}$, denoted $f(x, y)$, is integrable in x for each y and $\frac{\partial f}{\partial y}$ is integrable in x for each fixed value of y and is bounded: $|\frac{\partial f}{\partial y}(x, y)| \leq |g(x)|$ for some integrable function g . Then $\frac{d}{dy} \int f(x, y) dx = \int \frac{\partial f}{\partial y} dx$.*

Hypersurfaces

We will summarize some basic results about length of plane curves, area of surfaces, etc. A C^k surface $S \subset \mathbb{R}^3$ is a set of points so that, near each point $(x, y, z) \in S$, the points of S form the graph of a C^k function, say $x = f(y, z)$ or $y = g(x, z)$ or $z = h(x, y)$. (For example, the sphere $x^2 + y^2 + z^2 = 1$: the top is the graph of $z = \sqrt{1 - x^2 - y^2}$, while the bottom is the graph of $z = -\sqrt{1 - x^2 - y^2}$, and the right half is the graph of $x = \sqrt{1 - y^2 - z^2}$, etc.) For simplicity, let's assume that S is the graph of $z = h(x, y)$ over some open set D in the (x, y) -plane, and assume that $k \geq 1$. How can we define area? Let

S_ε be the set of points of distance at most $\varepsilon/2$ from S . Clearly we would like to have

$$\text{Vol}(S_\varepsilon) = \varepsilon \text{Area}(S) + o(\varepsilon).$$

It turns out that this equation forces us to measure area as: the area of the surface S is $\int_D \sqrt{1 + \partial_x h^2 + \partial_y h^2}$.

Similarly, a C^k hypersurface $S \subset \mathbb{R}^{n+1}$ is a set of points so that near each point we can write the surface as the graph of a C^k function, for example as $x_{n+1} = h(x_1, x_2, \dots, x_n)$ or similarly for some other coordinate. Suppose that S is actually the graph of such a function globally, say the graph of $h: D \rightarrow \mathbb{R}$ where $D \subset \mathbb{R}^n$. Any subset $X \subset S$ is then the graph of h over a subset $X_0 \subset D$. We then define the *hypersurface measure* of a measurable subset $X \subset S$ to be

$$\int_{X_0} \sqrt{1 + \|dh\|^2}.$$

It turns out that this is independent of how we choose to write S as the graph of a function, and the obvious analogue of equation 1 holds. Moreover, even if a C^k hypersurface S can only be written locally as a graph, we can add up local contributions from such integrals to get a globally defined hypersurface measure on measurable subsets of S . To define Lebesgue L^p and Hölder $C^{k,\alpha}$ functions on hypersurfaces, we follow precisely the same steps as we did before when we defined integration of functions on measurable sets in \mathbb{R}^n , but now using this hypersurface measure instead of outer measure.

A vector v is *tangent* to a hypersurface S at the point p if $v = x'(0)$ for some C^1 curve $x(t)$ so that $x(0) = p$. If $S \subset \mathbb{R}^{n+1}$ is C^1 then the tangent vectors to S at p form a hyperplane, i.e. a linear subspace of \mathbb{R}^{n+1} of dimension n , called the *tangent hyperplane*. A vector v is *perpendicular* to a hypersurface S at a particular point $p \in S$ if v is perpendicular to all of the tangent vectors to S at p . An *orientation* of a hypersurface is a continuous nowhere vanishing vector field n of unit vectors perpendicular to the hypersurface. By the implicit function theorem, if f is a C^k function and we let S be the set of points x at which both $f(x) = 0$ and $df(x) \neq 0$, then S is a C^k hypersurface; moreover S is orientable, since we can take $n = df / \|df\|$. The hypersurfaces of interest to us will be boundaries of domains. The boundary of a domain is orientable just when it is C^1 .

Theorem 1.8 (Divergence theorem). *If D is a domain with C^1 boundary and X is a compactly supported C^1 vector field defined on \bar{D} then*

$$\int_{\partial D} \langle X, n \rangle = \int_D \sum_i \partial_i X_i.$$

The left hand side is not mysterious: it measures how much X tends to stick out of the boundary of D . The right hand side is mysterious.

Weak convergence

A *Cauchy sequence* in a metric space X is a sequence x_1, x_2, \dots so that, no matter how close I want the elements of the sequence to be to each other, if I look out far enough down the sequence, any two of the elements will be no more than that close to each other. A metric space is *complete* if every Cauchy sequence converges. A *Banach space* is a normed vector space X which is complete, measuring distance from x to y as $\|x - y\|$. All of the spaces $L^p(U)$ on any open or closed set $U \subset \mathbb{R}^n$ are Banach spaces. A sequence $x_1, x_2, \dots \in X$ in a normed vector space *converges weakly* to an element $x \in X$ (called its *weak limit*) if, for any continuous linear function $f: X \rightarrow \mathbb{R}$, the numbers $f(x_1), f(x_2), \dots$ converge to $f(x)$. In any Banach space, a weakly convergent sequence x_1, x_2, \dots is bounded and has a unique weak limit $x = \lim x_j$ and $\|x\| \leq \liminf \|x_j\|$.

Theorem 1.9. *For any measurable subset $X \subset \mathbb{R}^n$ and any $1 < p < \infty$, every bounded sequence in $L^p(X)$ has a weakly convergent subsequence.*

The idea of the proof: pick a bounded sequence $h_1, h_2, \dots \in L^p(X)$. Fix any one continuous linear function $f: L^p(X) \rightarrow \mathbb{R}$; it is bounded on h_1, h_2, \dots , so you can pick a subsequence so that the values $f(h_1), f(h_2), \dots$ converge. Once you have done this for one linear function f , you can repeat the process for any finite number of such functions. With a little analysis, you can even do it for an infinite sequence of continuous linear functions $f_1, f_2, \dots: L^p(X) \rightarrow \mathbb{R}$. But all continuous linear functions on $L^p(X)$ have the form $f(h) = \int h\bar{g}$ for a unique $g \in L^q(\mathbb{R}^n)$, and there is a countable dense subset of $L^q(\mathbb{R}^n)$.

Chapter 2

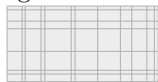
Approximation and Convolution

Convolution is a process of “smearing out” a function, which is often used to make smooth approximations to rough functions. We don’t have a clear intuition for rough functions. To prove a statement about rough functions, often we only need to prove it for smooth functions and invoke some type of continuity argument.

Approximating integrable functions

Theorem 2.1. *Every Riemann integrable function on any compact set is integrable with integral equal to the limit of Riemann sums.*

Proof. Take a Riemann integrable function f on a compact set $K \subset \mathbb{R}^n$. Cover



K in a grid of boxes like: For each point $x \in K$, let $L(x)$ to be the infimum of $f(y)$ over all y in the grid box containing x , and $U(x)$ be the supremum. So $L < f < U$. By definition, L and U are integrable functions, and $\int L$ is the lower Riemann sum for this grid, and $\int U$ the upper. Since f is Riemann integrable, L and U pointwise approach f as we refine the grid to a finer mesh. Clearly L increases as we refine the grid, and U decreases, so we can apply the dominated convergence theorem to prove that f is integrable with $\int f$ the limit of the Riemann sums. \square

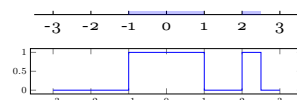
The *indicator function* of a set X is the function

$$1_X(x) = \begin{cases} 1 & \text{if } x \in X, \\ 0 & \text{otherwise.} \end{cases}$$

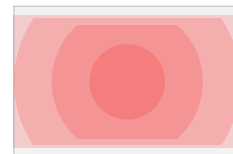
A *New York function* is a finite sum $\sum a_j 1_{X_j}$ where the sets X_j are boxes.

Lemma 2.2. *Pick p with $1 \leq p < \infty$ and an open set $U \subset \mathbb{R}^n$. Every function $f \in L^p(U)$ is the limit of (1) a sequence of test functions supported in U , and (2) a sequence of New York functions supported in U . In other words, the test functions are dense in $L^p(U)$, as are the New York functions.*

Proof. The set of test functions is a linear subspace of L^p . Therefore the set of L^p functions which arise as limits of such functions is also a linear subspace. (Clearly the same argument works for New York functions.) Take any $f \in L^p$.



A set X and its indicator function 1_X



An infinitely wide horizontal strip U and various compact sets X_k “approximating” it

We can assume that f is real-valued. Pick a point $x_0 \in U$. For each integer $k > 0$, let X_k be the set of points $x \in U$ so that $d(x, x_0) \leq k$ and so that $d(x, \partial U) \geq 1/k$. Clearly X_k is compact. Let

$$f_k(x) = \begin{cases} f(x) & \text{if } x \in X_k, \\ 0 & \text{otherwise.} \end{cases}$$

By the dominated convergence theorem, $f_k \rightarrow f$ in L^p . So it suffices to prove the theorem for functions $f \in L^p$ with compact support, so assume f has compact support.

Next we make a discrete approximation to f , dividing a large range of values $-k^2 < y < k^2$ into small steps of size $1/k$ and rounding off f to the nearest y value that lies at one of those steps. By $\lfloor t \rfloor$ denote the largest integer less than or equal to a real number t . Let

$$f_k(x) = \begin{cases} \frac{\lfloor kf(x) \rfloor}{k}, & \text{if } |f(x)| < k^2, \\ 0 & \text{otherwise.} \end{cases}$$

a finite linear combination of indicator functions of bounded measurable sets. Clearly $|f_k - f|^p \leq |f|^p$, so by the dominated convergence theorem, $f_k \rightarrow f$ in L^p -norm. It suffices to prove the result for indicator functions of bounded measurable sets $f = 1_X$.

Let W be a open set containing X so that the measure of the difference is as small as we like. Since X is bounded, we can take W to be bounded. Then clearly $1_W \rightarrow 1_X$ in L^p as the measure of the difference gets small. So it suffices to prove the result for indicator functions of bounded open sets $f = 1_W$.

Alternatively, let W be an open set containing $\mathbb{R}^n - X$ so that the difference has measure as small as we like, and let $Y = \mathbb{R}^n - W$. So $Y \subset X$ is a closed set and $X - Y$ has measure as small as we like. As we make that measure small, $1_Y \rightarrow 1_X$ in L^p . So it suffices to prove the result for $f = 1_Y$ the indicator function of a compact set.

Take a smooth function $h: \mathbb{R} \rightarrow \mathbb{R}$ so that $h(x) = 1$ if $x \leq 0$ and $h(x) = 0$ if $x \geq 1$. Let $d(x)$ be the distance from x to Y and let

$$f_k(x) = h(kd(x));$$

$|f_k - f|^p$ goes to zero pointwise and is bounded by $|f_1|^p$ so by dominated convergence $\|f_k - f\|_{L^1} \rightarrow 0$. So test functions are dense in L^p .

As for New York functions, as above it suffices to prove the result for indicator functions of bounded open sets, say $f = 1_U$. Draw a very large box, and cut it into a very fine mesh. Let X be the union of all of the grid boxes of this mesh that live entirely inside U . Every point of U lies in some such a grid box, for some fine enough mesh inscribed into a large enough box, so as we refine the mesh and make the box larger, 1_X will approach 1_U pointwise, and so in L^p (by the dominated convergence theorem). \square

2.1 Suppose that $U, U^* \subset \mathbb{R}^n$ are open sets and $F: U \rightarrow U^*$ is a C^1 map with C^1 inverse. For any $f \in L^1(U^*)$, prove that $x \mapsto f(F(x)) |\det F'(x)| \in L^1(U)$ and Prove that

$$\int_U f(F(x)) |\det F'(x)| = \int_{U^*} f.$$

Hint: you already know this is true for continuous functions with compact support, using the Riemann integral, from your earlier analysis courses.

Theorem 2.3 (Continuity of translation). *Suppose that $1 \leq p < \infty$. For any $f \in L^p$, $\|f(x+y) - f(x)\|_{L^p} \rightarrow 0$ as $y \rightarrow 0$.*

Proof. The set of functions f for which this result is true is a linear subspace of L^p : just imagine adding or scaling. Suppose that the result is true for f in some dense linear subspace of L^p . Then for any f , take some sequence f_1, f_2, \dots in that subspace so that $f_j \rightarrow f$ in L^p and

$$\begin{aligned} & \|f(x+y) - f(x)\|_{L^p} \\ & \leq \|f(x+y) - f_j(x+y)\|_{L^p} + \|f_j(x+y) - f_j(x)\|_{L^p} + \|f_j(x) - f(x)\|_{L^p}; \end{aligned}$$

take $j \rightarrow \infty$. So it suffices to prove the result for the indicator function of a box, for which it is easy to picture and check explicitly. \square

Convolution

If $f \in L^p(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$ with $\frac{1}{p} + \frac{1}{q} = 1$, let

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

The function g appears in the integral translated, and Lebesgue norms are translation invariant, so by Hölder's inequality

$$\|f * g\|_{L^\infty} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Similarly if we let $p = 1$ and $q = \infty$.

Imagine listening to a conversation through a wall. We let $f(t)$ be the sound (i.e. density of air approaching the wall, at time t). Then the sound you hear through the wall is not $f(t)$, but some muffling of $f(t)$, because the sound bounces around inside the wall for a little while. So at time t you hear an average of values of f at times earlier than t , say

$$\int_{\mathbb{R}} f(s) g(t-s) ds,$$

where $g(t-s)$ represents how much signal gets through at time t from the sound made at time s . So $g(t)$ represents how much signal gets through at time t from the sound made at time $s = 0$. This $f * g$ represents f “smeared out” by averaging against g .

Lemma 2.4. $f * g = g * f$

Proof. Change variable by $z = x - y$: $\int_{\mathbb{R}^n} f(y)g(x-y) dy = \int_{\mathbb{R}^n} f(x-z)g(z) dz$. \square

By the dominated convergence theorem, if f is continuous and bounded and $g \in L^1$ then $f * g$ is continuous.

2.2 If g is C^1 with bounded derivative and $f \in L^1$, prove that the dominated convergence theorem allows us to differentiate under the integral sign to reveal that

$$\partial_i (f * g) = f * \partial_i g.$$

Similarly, if $g \in C^\infty$ and all derivatives of g are bounded, then

$$f * \partial^a g = \partial^a f * g,$$

so that $f * g \in C^\infty$ with all derivatives bounded.

Lemma 2.5. If $f, g \in L^1$ then $f * g \in L^1$ and $\|f * g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1}$.

Proof. By Fubini's theorem (theorem 1.3),

$$\begin{aligned} \|f * g\|_{L^1} &\leq \int \left(\int |f(y)g(x-y)| dy \right) dx, \\ &= \int \left(\int |f(y)g(x-y)| dx \right) dy, \\ &= \int |f(y)| \left(\int |g(x-y)| dx \right) dy, \\ &= \int |f(y)| \|g\|_{L^1} dy, \\ &= \|f\|_{L^1} \|g\|_{L^1}. \end{aligned}$$

\square

2.3 Suppose that $f, g, h \in L^1$. Let $k(x) = \bar{h}(-x)$. Use Fubini's theorem to prove that $\langle f * h, g \rangle = \langle f, k * g \rangle$.

2.4 Prove that f and g are integrable, then $\int f * g = \int f \int g$.

Theorem 2.6 (Hausdorff–Young inequality). If $f \in L^1$ and $g \in L^p$, then $f * g \in L^p$ and $\|f * g\|_{L^p} \leq \|f\|_{L^1} \|g\|_{L^p}$.

Proof. Since $|f * g| \leq |f| * |g|$, we can assume that f and g are nonnegative. We can also assume that $p > 1$, since the result for $p = 1$ is our previous lemma. If

$p = \infty$,

$$\begin{aligned} f * g(x) &\leq \int \|f\|_{L^\infty} g(x-y) dy, \\ &= \|f\|_{L^\infty} \int g(x-y) dy, \\ &= \|f\|_{L^\infty} \|g\|_{L^1}. \end{aligned}$$

So we can assume that $1 < p < \infty$. Take q so that $\frac{1}{p} + \frac{1}{q} = 1$. Then

$$f * g(x) = \int f(y) g(x-y)^{1/p} g(x-y)^{1/q} dy,$$

to which we apply Hölder's inequality

$$\begin{aligned} &\leq \left(\int f(y)^p g(x-y) dy \right)^{1/p} \left(\int g(x-y) dy \right)^{1/q}, \\ &= ((f^p) * g)(x)^{1/p} \|g\|_{L^1}^{1/q}. \end{aligned}$$

In this series of inequalities, take the first expression and the last each to the power of p :

$$(f * g(x))^p \leq f^p * g(x) \|g\|_{L^1}^{p/q}.$$

Integrate

$$\begin{aligned} \|f * g\|_{L^p}^p &\leq \|g\|_{L^1}^{p/q} \int f^p * g, \\ &= \|g\|_{L^1}^{p/q} \int f^p \int g \\ &= \|f\|_{L^p}^p \|g\|_{L^1}^{1+p/q}, \\ &= \|f\|_{L^p}^p \|g\|_{L^1}^p. \end{aligned}$$

Take p -th roots. □

Approximation of the identity

Theorem 2.7. Suppose that $f \in L^1(\mathbb{R}^n)$ and that $\int f = 1$. Let $f_\varepsilon(x) = \varepsilon^{-n} f(\frac{x}{\varepsilon})$. For any p with $1 \leq p < \infty$, and for any $g \in L^p(\mathbb{R}^n)$, let $g_\varepsilon = f_\varepsilon * g$. Then $g_\varepsilon \rightarrow g$ in L^p as $\varepsilon \rightarrow 0$. The same is true in L^∞ if g is continuous.

Proof. Suppose that $\frac{1}{p} + \frac{1}{q} = 1$ and as usual if $p = \infty$ then set $q = 1$, and if $p = 1$ then set $q = \infty$. Writing f as a difference of nonnegative integrable functions, it suffices to prove the result for $f \geq 0$. Start with $1 \leq p < \infty$. Clearly

$$g(x) = \int g(x) f_\varepsilon(y) dy.$$

Therefore

$$\begin{aligned} |g_\varepsilon(x) - g(x)| &= \left| \int (g(x-y) - g(x)) f_\varepsilon(y) dy \right|, \\ &= \left| \int (g(x-\varepsilon y) - g(x)) f(y) dy \right|, \end{aligned}$$

and factoring f into two pieces (and if $p = 1$, let $1/q = 0$)

$$\leq \int |g(x-\varepsilon y) - g(x)| f(y)^{1/p} f(y)^{1/q} dy.$$

Apply Hölder's inequality, and then raise both sides to the p -th power and integrate:

$$\begin{aligned} \int |g_\varepsilon - g|^p &\leq \int \left\{ \left(\int |g(x-\varepsilon y) - g(x)|^p f(y) dy \right) \left(\int f(y) dy \right)^{p/q} \right\} dx, \\ &= \iint |g(x-\varepsilon y) - g(x)|^p f(y) dy dx, \\ &= \iint |g(x-\varepsilon y) - g(x)|^p dx f(y) dy. \end{aligned}$$

But $\int |g(x-\varepsilon y) - g(x)|^p dx \rightarrow 0$ pointwise in y as $\varepsilon \rightarrow 0$ by continuity of translation (theorem 2.3 on page 13), and is bounded by $2 \|g\|_{L^p}^p$ as a function of y , so the dominated convergence theorem says that $\int |g_\varepsilon - g|^p \rightarrow 0$.

Next, try $p = \infty$, and assume g continuous. As above,

$$|g_\varepsilon(x) - g(x)| = \int |g(x-\varepsilon y) - g(x)| f(y) dy.$$

Pick a large closed ball \bar{B} . The part of the integral occurring over \bar{B} gets small, because $g(x-\varepsilon y)$ converges to g uniformly on \bar{B} by continuity. The part of the integral away from there gets small because g is uniformly bounded, and f is integrable so has small integral on $\mathbb{R}^n - \bar{B}$ for large enough \bar{B} . \square

The *Gaussian* is the function $f(x) = e^{-\|x\|^2}$, also called a *bell curve*. Any translate or rescaling of this function will also be called a *Gaussian*, i.e. the functions $ae^{-b\|x-x_0\|^2}$, for $b > 0$ and $x_0 \in \mathbb{R}^n$.

Lemma 2.8. *Every $h \in L^p(\mathbb{R}^n)$, if $1 \leq p < \infty$, is the limit of a sequence of functions h_1, h_2, \dots where each h_j is a finite sum of Gaussians. In other words, the Gaussians span a dense linear subspace of L^p .*

Proof. Rescale a Gaussian f to have $\int f = 1$. We can assume that h is a test function, because such functions are dense in L^p by lemma 2.2 on page 11. Then for any $h \in L^p$, $f_\varepsilon * h \rightarrow h$ in L^p . This convolution is

$$f_\varepsilon * h = \int f_\varepsilon(x-y) h(y) dy.$$

Since h is continuous with compact support, we can approximate this integral with a Riemann sum: a finite sum of Gaussians. \square

Theorem 2.9. *Suppose that $U \subset \mathbb{R}^n$ is an open set and $g \in L^p_{loc}(U)$. Extend g to be 0 outside of U , so $g \in L^p(\mathbb{R}^n)$. Let f be a test function with $\int f = 1$ vanishing outside the unit ball, let $f_\varepsilon(x) = \varepsilon^{-n}f(x/\varepsilon)$, and let*

$$g_\varepsilon = f_\varepsilon * g.$$

Then g_ε is C^∞ ; if U is bounded then g_ε is a test function. Moreover $g_\varepsilon \rightarrow g$ in $L^p_{loc}(U)$ as $\varepsilon \rightarrow 0^+$.

Proof. The proof is the same as theorem 2.7 on page 15 above. \square

2.5 Suppose that $K \subset \mathbb{R}^n$ is a compact set and $U \subset \mathbb{R}^n$ is an open set containing K . Prove that there is a test function f supported in U so that $0 \leq f \leq 1$ and $f = 1$ at every point of K .

Chapter 3

Sobolev spaces

Sobolev spaces are spaces of functions whose derivatives up to some order live in L^p . They are the right place to look for solutions to many differential equations.

Weak derivatives

Many functions don't have derivatives at some points. We need a weaker notion of derivative, which pays less attention to poorly behaved points. First, suppose we have a differentiable function: if f is a C^1 function on \mathbb{R} , then for any test function ϕ ,

$$\int_{-\infty}^{\infty} f' \phi = - \int_{-\infty}^{\infty} f \phi',$$

by integration by parts, and using the fact that ϕ vanishes outside some interval. Similarly, if f is a C^1 function on \mathbb{R}^n , then

$$\int (\partial_i f) \phi = - \int f (\partial_i \phi)$$

and more generally if f is C^k , then

$$\int (\partial^a f) \phi = (-1)^{|a|} \int f \partial^a \phi.$$

Consider the function $f(x) = |x|^\alpha$, for some positive constant α , $0 < \alpha < 1$. Then $f(x)$ is not differentiable at $x = 0$ (the graph of f is a cusp), but still if we differentiate away from 0, we find

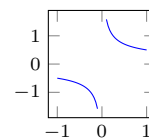
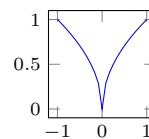
$$f'(x) = \alpha \operatorname{sgn}(x) |x|^{\alpha-1}.$$

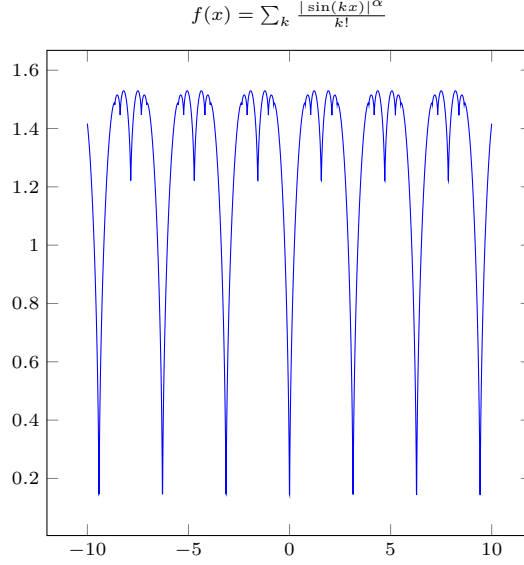
Note that f' is a locally integrable function, but if $\alpha < 1$ then f' has an infinite spike at $x = 0$.

Nevertheless, it is easy to check that integration by parts works fine:

$$\int_{-\infty}^{\infty} f' \phi = - \int_{-\infty}^{\infty} f \phi',$$

for any test function ϕ .





We say that a locally integrable function g is a *weak derivative* of a locally integrable function f , symbolically $g = \partial^a f$, if

$$\int f \partial^a \phi = (-1)^a \int g \phi,$$

for any test function ϕ .

3.1 Prove that f only has one such weak derivative $\partial^a f$, i.e. any two locally integrable functions g and h so that $\int g \partial^a \phi = \int h \partial^a \phi$ for every test function ϕ must satisfy $g = h$ (except perhaps on a set of measure zero, but of course we identify such functions anyway).

Derivatives in the usual sense will be called *strong derivatives*. A more serious example: pick a constant $0 < \alpha < 1$ and let

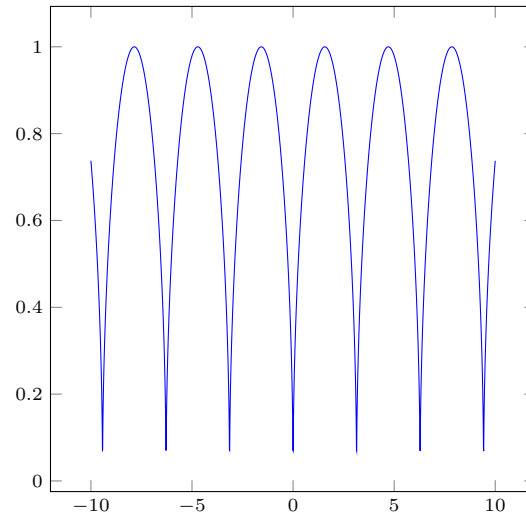
$$f(x) = \sum_k \frac{|\sin(kx)|^\alpha}{k!}.$$

One easily checks that $f(x)$ has weak derivative

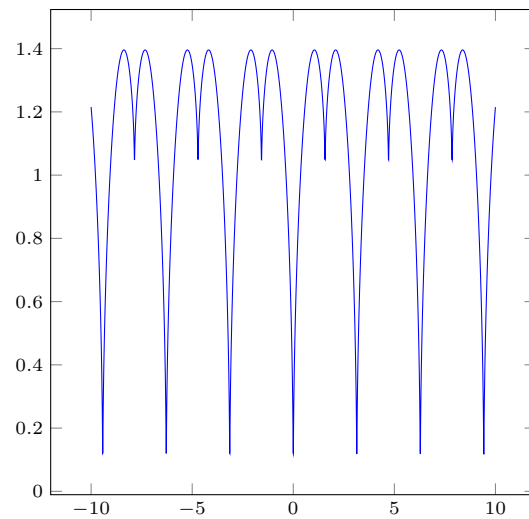
$$f'(x) = \sum_k \frac{\operatorname{sgn}(x) \cos(kx) |\sin(kx)|^{\alpha-1}}{(k-1)!}.$$

This function $f'(x)$ has a spike going to $\pm\infty$ at every point x where x is any rational multiple of π , but f' is well defined away from those points and locally integrable. Those points form a dense but measure zero set. A picture of f' looks like spikes going to $\pm\infty$ all over the place, and f is not differentiable (in the usual, strong, sense) anywhere.

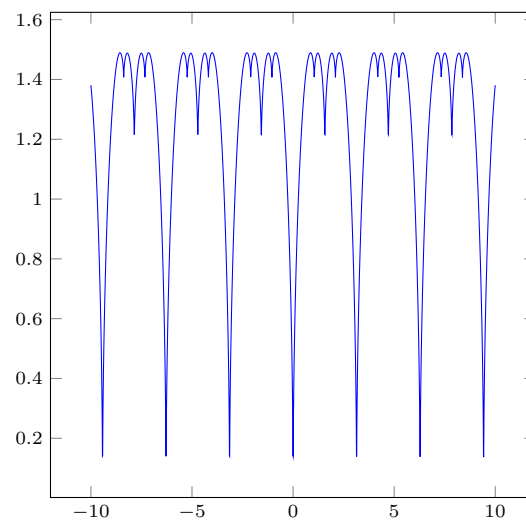
The first term in the sum for $f(x): |\sin(x)|^\alpha$



The first 2 terms in the sum for $f(x): |\sin(x)|^\alpha + \frac{1}{2} |\sin(x/2)|^\alpha$



The first 3 terms in the sum for $f(x)$



3.2 Prove that the *Heaviside function*

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1, & \text{if } x > 0, \end{cases}$$

does not have a weak derivative in L^1_{loc} .

3.3 Let $f_1: [0, 1] \rightarrow \mathbb{R}$ be the piecewise linear function with values $f(0) = 1, f(1/3) = 1, f(1/2) = 0, f(2/3) = 1, f(1) = 1$ and linear in between each of these points. Inductively, let

$$f_{k+1}(x) = \begin{cases} f_k(3x), & \text{if } 0 \leq x \leq 1/3, \\ f_k(x), & \text{if } 1/3 \leq x \leq 2/3, \\ f_k(3(x - 2/3)), & \text{if } 2/3 \leq x \leq 1. \end{cases}$$

Draw f_1, f_2, f_3 . Let $f(x) = \lim_{k \rightarrow \infty} f_k(x)$. Prove that $f(x)$ has a weak derivative in $L^\infty([0, 1])$ which is discontinuous on an uncountable set of points.

Weak derivatives sound complicated. Avoid them: first prove that smooth functions are dense in whatever function space we want to work in, and from then on we only need to prove theorems about smooth functions, using usual derivatives, and (if the statements in the theorems behave well under taking limits) the proof is done for the whole function space.

Theorem 3.1. *Suppose that $U \subset \mathbb{R}^n$ is an open set. Suppose that $g \in L^1_{\text{loc}}(U)$ has a weak derivative $\partial^a g \in L^1_{\text{loc}}(U)$. Suppose that f is a test function with $\int f = 1$ and let $f_\varepsilon(x) = \varepsilon^{-n} f(x/\varepsilon)$. Then $f_\varepsilon * g \in C^\infty(U)$ and $\partial^a (f_\varepsilon * g) = (\partial^a f_\varepsilon) * g = f_\varepsilon * \partial^a g$. As $\varepsilon \rightarrow 0^+$, $\partial^a (f_\varepsilon * g) \rightarrow \partial^a g$ in $L^1_{\text{loc}}(U)$. In particular, smooth functions are dense in the space of locally integrable functions with any number of prescribed weak derivatives.*

Proof. Theorem 2.9 on page 17 tells us that $f_\varepsilon * g \in C^\infty(U)$. Differentiation under the integral sign shows that $\partial^a (f_\varepsilon * g) = (\partial^a f_\varepsilon) * g$. Note that $\partial_{y_i} f(x - y) = -\partial_{x_i} f(x - y)$. By induction, $\partial_y^a f(x - y) = (-1)^{|a|} \partial_x^a f(x - y)$. Therefore

$$\begin{aligned} (\partial^a f_\varepsilon) * g(x) &= \int \partial_x^a f_\varepsilon(x - y) g(y) dy, \\ &= (-1)^{|a|} \int \partial_y^a f_\varepsilon(x - y) g(y) dy, \\ &= \int f_\varepsilon(x - y) \partial_y^a g(y) dy, \\ &= f_\varepsilon * \partial^a g. \end{aligned}$$

Theorem 2.9 on page 17 now applies to $f_\varepsilon * \partial^a g$. □

We want the freedom to approach a function with weak derivative along a sequence of smooth functions, not necessarily just by convolution.

Theorem 3.2. Suppose that f_1, f_2, \dots is a sequence of smooth functions converging in $L^1_{loc}(U)$ and that $\partial^a f_j$ also converges in $L^1_{loc}(U)$. Then $\lim_j \partial^a f_j = \partial^a \lim_j f_j$.

Proof. Let $f = \lim_j f_j$ and $g = \lim_j \partial^a f_j$. For any test function ϕ , the compact support of ϕ ensures that ϕg and $\phi \partial^a f_j$ lie in $L^1(U)$, and the Hölder inequality applied to $\int \phi(g - \partial^a f_j)$ ensures that

$$\begin{aligned} \int \phi g &= \lim_j \int \phi(\partial^a f_j), \\ &= (-1)^{|a|} \lim_j \int f_j \partial^a \phi, \end{aligned}$$

but then Hölder again gives us

$$= (-1)^{|a|} \int f \partial^a \phi.$$

□

Sobolev spaces

If $U \subset \mathbb{R}^n$ is an open set, the *Sobolev space* $L^p_k(U)$ to be the set of all functions $f \in L^p$ so that f has weak derivative $\partial^a f \in L^p$ for any a with $|a| \leq k$. The *Sobolev norm* of a function $f \in L^p_k(U)$ is

$$\|f\|_{L^p_k} = \left(\sum_{|a| \leq k} \|\partial^a f\|_{L^p}^p \right)^{1/p}.$$

Clearly $\dots \subset L^p_2 \subset L^p_1 \subset L^p_0 = L^p$. If U is a bounded open set, then $q \geq p$ implies $L^q(U) \subset L^p(U)$, and so as we raise either p or k , $L^p_k(U)$ gets smaller, a more restrictive Sobolev space. It follows immediately from the completeness of the L^p spaces and the Hölder inequality that each Sobolev space is complete in its norm.

3.4 For each real number α , what Sobolev spaces does $|x|^\alpha e^{-x^2}$ belong to?

In the study of partial differential equations we are most often faced with a sequence of functions in a Sobolev space, which might only converge to a function in another, less restrictive Sobolev space. The two main theorems about Sobolev spaces tells us (1) when a Sobolev function is continuous (or more generally, when it is C^k) and (2) when a sequence of functions in one Sobolev space must converge to a function, but perhaps in different, less restrictive Sobolev space.

3.5 Use theorem 1.9 on page 9 to prove that every bounded sequence in $L^p_k(U)$ has a weakly convergent subsequence.

Density of the test functions

Lemma 3.3. *If $1 \leq p < \infty$ then $C^\infty(\mathbb{R}^n) \cap L_k^p(\mathbb{R}^n)$ is dense in $L_k^p(\mathbb{R}^n)$.*

Proof. Take a function $g \in L_k^p(\mathbb{R}^n)$ and a test function f with $\int f = 1$. Let $f_\varepsilon(x) = \varepsilon^{-n} f(x/\varepsilon)$ and $g_\varepsilon = f_\varepsilon * g$. Theorem 3.1 on page 22 shows that g_ε is smooth and, as $\varepsilon \rightarrow 0$, the various derivatives of g_ε will converge in L_{loc}^1 to the corresponding weak derivatives of g . Theorem 2.7 on page 15 proves that they converge in L^p . \square

Theorem 3.4. *If $1 \leq p < \infty$ then the test functions are dense in $L_k^p(\mathbb{R}^n)$.*

Proof. By lemma 3.3, it suffices to prove density of the test functions among $C^\infty(\mathbb{R}^n) \cap L_k^p(\mathbb{R}^n)$. Pick a smooth function $g \in C^\infty(\mathbb{R}^n) \cap L_k^p(\mathbb{R}^n)$ and a test function f so that $0 \leq f \leq 1$ and f equals 1 near the origin. Let $f_\varepsilon(x) = f(\varepsilon x)$ and $g_\varepsilon = f_\varepsilon g$. So $g_\varepsilon(x) = g(x)$ in the ball of radius ε^{-1} . By the dominated convergence theorem, $g_\varepsilon \rightarrow g$ in L^p . By the product rule,

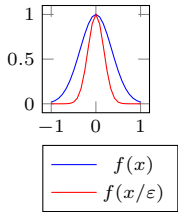
$$\partial_i g_\varepsilon = f_\varepsilon \partial_i g + \varepsilon \partial_i f(\varepsilon x) g.$$

As $\varepsilon \rightarrow 0$, $f_\varepsilon \partial_i g \rightarrow \partial_i g$ by the dominated convergence theorem. We can apply Hölder's inequality to $\varepsilon \partial_i f(\varepsilon x) g$, because the derivatives of f have the same height no matter what ε , so this vanishes with ε . By induction, the same tricks work for derivatives of all orders. \square

3.6 Prove that the smooth functions are not dense in the Hölder space $C^{0,1}(\mathbb{R})$ by: (1) showing that $f(x) = |x|$ belongs to this space, but that (2) if g is any function belonging to this space and g is differentiable at 0, say with $g'(0) \geq 0$, then

$$\frac{|(f(x) + g(x)) - (f(0) + g(0))|}{|x|}$$

has limit as $x \rightarrow 0^-$ given by $g'(0) + 1$, while if $g'(0) \leq 0$ then as $x \rightarrow 0^+$ it has limit $|g'(0)| + 1$. Carry out a similar trick to prove that smooth functions are not dense in $C^{0,\alpha}$ for $0 < \alpha < 1$.



Sensitivity to small bumps and high frequencies

Pick any function f and rescale it: define $f_\varepsilon(x) = f(x/\varepsilon)$. More generally, suppose that we have some operation T_ε taking some functions f to functions $T_\varepsilon f$, where the operation depends on some parameters ε in some space \mathbb{R}^k . Write $T_\varepsilon f$ as f_ε , a parameterized family of functions. Suppose that X is some space of functions, equipped with some norm. Pick some $f \in X$ for which f_ε is defined and lies in X for arbitrarily small values of ε . Imagine that we find that

$$\|f_\varepsilon\|_X = \frac{a(f) + o(1)}{|\varepsilon|^\beta} \text{ as } |\varepsilon| \rightarrow 0.$$

Suppose that this equation persists with the same β for any such f , and there is some f for which $a(f) \neq 0$. The number β is the *sensitivity* to the operation T_ε . Some function spaces X won't have a defined sensitivity to that operation, because we can't carry out the operation on any functions and still stay in X , or because there is no such number β .

If $T_\varepsilon f(x) = f(x/\varepsilon)$, the associated sensitivity is the *sensitivity to small bumps* of X , which we denote by σX . By the chain rule, $\partial_i f_\varepsilon(x) = \partial_i f(\varepsilon x)/\varepsilon$, derivatives scale by factors of ε^{-1} : $\sigma C^k = k$. Similarly,

$$\|f_\varepsilon\|_{C^{k,\alpha}} = \frac{\|f\|_{C^{k,\alpha}} + o(1)}{\varepsilon^{k+\alpha}}$$

so $\sigma C^{k,\alpha} = k + \alpha$. Lets find the sensitivity to small bumps of each Sobolev space. Our integrands are

$$|\partial^a f_\varepsilon(x)|^p = \frac{|\partial^a f(x/\varepsilon)|^p}{\varepsilon^{p|a|}}.$$

Integrating, when we rescale x we rescale all of \mathbb{R}^n , so rescale volumes by ε^n :

$$\int |\partial^a f_\varepsilon(x)|^p = \frac{\int |\partial^a f|^p}{\varepsilon^{p|a|-n}}.$$

Taking p -th roots,

$$\left(\int |\partial^a f_\varepsilon(x)|^p \right)^{1/p} = \frac{\left(\int |\partial^a f|^p \right)^{1/p}}{\varepsilon^{|a|-n/p}}.$$

As $\varepsilon \rightarrow 0$, ignoring the lower order terms, we get

$$\|f_\varepsilon\|_{L_k^p} = \frac{\|f\|_{L_k^p} + o(1)}{\varepsilon^{k-n/p}}.$$

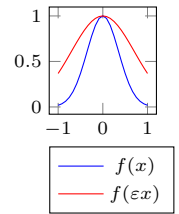
So $\sigma L_k^p(\mathbb{R}^n) = k - \frac{n}{p}$. This roughly tells us to expect that functions in L_k^p have k weak derivatives but only $k - \frac{n}{p}$ strong derivatives.

Pick any function f and rescale it the other way: define $T_\varepsilon f(x) = f(\varepsilon x)$. For functions f defined in all of \mathbb{R}^n , this defines a sensitivity. The associated sensitivity is the *sensitivity to large humps* of X , which we denote by λX . By the chain rule, $\partial_i f_\varepsilon(x) = \varepsilon \partial_i f(\varepsilon x)$, derivatives scale by factors of ε , so the zeroth derivative contributes the most: $\lambda C^k = 0$. Similarly,

$$\|f_\varepsilon\|_{C^{k,\alpha}} = \varepsilon^\alpha (\|f\|_{C^{k,\alpha}} + o(1))$$

so $\lambda C^{k,\alpha} = -\alpha$.

3.7 Explain why, in any bounded domain containing the origin, the “sensitivity to large humps” doesn't actually define a sensitivity on Sobolev spaces. Hint: try the bounded domain $0 \leq x \leq 1$ in \mathbb{R} . Nonetheless we persist to use the notation λX even in bounded domains.



But $\lambda L_k^p(\mathbb{R}^n) = \frac{n}{p}$, so we use the notation λX to mean $\frac{n}{p}$ for any Sobolev space $X = L_k^p$.

If we pick a vector $\xi \in \mathbb{R}^n$ and let

$$f_\varepsilon(x) = e^{2\pi i \langle \xi, x \rangle / \varepsilon} f(x),$$

the associated sensitivity is the *sensitivity to high frequencies* of X , which we denote $\gamma = \phi X$.

3.8 Prove for functions in \mathbb{R}^n :

	σ	λ	ϕ
L_k^p	$k - \frac{n}{p}$	$\frac{n}{p}$	k
C^k	k	0	k
$C^{k,\alpha}$	$k + \alpha$	$-\alpha$	k

and consequently $\sigma + \lambda = \phi$ for all of these function spaces.

If X and Y are function spaces (i.e. vector spaces of functions), equipped with norms, X is *embedded* in Y if X is a linear subspace of Y and that there is a constant C so that, for every $f \in X$, $\|f\|_Y \leq C \|f\|_X$. The minimum possible value of C is the *best constant* of the embedding. If furthermore every bounded sequence in X has a convergent subsequence in Y , we say that $X \subset Y$ is a *compactly embedded subspace*.

3.9 Prove $L^\infty([0, 1]) \subset L^1([0, 1])$ is embedded, with best constant $C = 1$.

3.10 If $X \subset \mathbb{R}^n$ is a set of finite volume, use the Hölder inequality to prove that $L^q(X) \subset L^p(X)$ is embedded, for $1 \leq p \leq q$. Find the best constant in terms of the volume of X .

We expect that if Y is less sensitive than X , then Y tolerates more functions than X does, i.e. is willing to contain more functions.

Lemma 3.5. *If $X \subset Y$ is embedded and both of X and Y have sensitivities to some family of operations T_ε (for example: small bumps, large humps or high frequencies), then the sensitivities of Y are less than or equal to those of X .*

Proof. If $\sigma X < \sigma Y$, then we can scale $\varepsilon \rightarrow 0$ and the ratio $\|f_\varepsilon\|_Y / \|f_\varepsilon\|_X$ grows like a negative power of ε . Embedding of $X \subset Y$ says that $\|f_\varepsilon\|_Y \leq C \|f_\varepsilon\|_X$, i.e. $\|f_\varepsilon\|_Y / \|f_\varepsilon\|_X \leq C$, not like a negative power of ε . The same argument works for the other sensitivities. \square

In a bounded domain, functions in L^p , as we increase p , are more tightly controlled with thinner spikes. Each L^p space is embedded in every $L^{p-\varepsilon}$ space for $\varepsilon > 0$: the smaller p gets, the wider the spikes can be.

Imagine that we want to see if some Sobolev space $X = L_{k_0}^{p_0}(U)$ is embedded inside some other Sobolev space $Y = L_{k_1}^{p_1}(U)$, or Hölder space $Y = C^{k_1, \alpha_1}(\bar{U})$. Roughly speaking (although this is not quite true), if the sensitivities of X are bigger than those of Y , then we expect that X is embedded in Y . If \bar{U} is a bounded domain, and if we are willing to decrease k_0 to some smaller value k_1 , we might be able to increase p_0 to some slightly larger value p_1 and still obtain an embedding. Note that this goes against the grain, since increasing p_0 to p_1 is *not* an embedding of L^p spaces. In other words, we trade off: we lose derivatives ($k_0 > k_1$) but gain control on the spikes ($p_0 < p_1$).

Theorem 3.6 (The Sobolev embedding theorem). *Suppose that $\bar{U} \subset \mathbb{R}^n$ is a compact domain with C^1 boundary, and X is a Sobolev space with $\lambda X > 0$ and Y is a Sobolev or Hölder space or a space of functions with bounded derivatives, i.e. $L_k^p(U)$ with $1 \leq p \leq \infty$ or $C^{k, \alpha}(\bar{U})$ with $0 < \alpha < 1$ or $C_b^k(\bar{U})$. If the sensitivities of Y are all less than or equal to those of X and*

- a. $\lambda Y \geq 0$ and
 - 1. $\sigma X > \phi Y$ or
 - 2. $\sigma X = \phi Y$ and $\lambda X = n$ or
 - 3. $\lambda Y > 0$ and
 - i) $\sigma X = \phi Y$ or
 - ii) $\phi X > \phi Y$ or
 - iii) $\lambda X = n$

or

- b. $\lambda Y < 0$ and $\lambda X > 0$ and $\sigma X > \phi Y > \sigma X - 1$

then $X \subset Y$ is an embedded subspace.

Theorem 3.7 (The Kondrashov–Rellich compactness theorem). *If the \geq signs governing the sensitivities in the Sobolev embedding theorem are $>$ signs, then the embedding is compact.*

3.11 Suppose that $L_{k_0}^{p_0}(U) \subset L_{k_1}^{p_1}(U)$ is an embedding; prove that $L_{k_0+1}^{p_0}(U) \subset L_{k_1+1}^{p_1}(U)$ is too.

3.12 Suppose that $L_{k_0}^{p_0}(U) \subset C^{k_1, \alpha_1}(U)$ is an embedding; prove that $L_{k_0+1}^{p_0}(U) \subset C^{k_1+1, \alpha_1}(U)$ is too.

The fundamental theorem of calculus in one variable

Lemma 3.8. *Suppose that $f \in L_1^1(\mathbb{R})$. Then f is bounded and continuous and $f(x) \rightarrow 0$ as $x \rightarrow \infty$ and as $x \rightarrow -\infty$. Moreover, $\|f\|_{C^0} \leq \|f\|_{L^1} \leq \|f\|_{L_1^1}$, so $L_1^1(\mathbb{R}) \subset C_b^0(\mathbb{R})$ is an embedded subspace.*

Proof. Assume that f is a test function. By the fundamental theorem of calculus

$$f(x_1) - f(x_0) = \int_{x_0}^{x_1} f'(u) du.$$

Therefore

$$\begin{aligned} |f(x_1) - f(x_0)| &\leq \int_{x_0}^{x_1} |f'(u)| du, \\ &\leq \int_{-\infty}^{\infty} |f'(u)| du, \\ &= \|f'\|_{L^1}. \end{aligned}$$

If we take x_0 outside of the support of f , we find $|f(x)| \leq \|f'\|_{L^1}$. Taking supremum,

$$\|f\|_{C^0} \leq \|f'\|_{L^1} \leq \|f\|_{L^1} + \|f'\|_{L^1} = \|f\|_{L^1_1}.$$

Take limits of test functions; we leave the reader to prove that the test functions are dense among the bounded continuous functions that vanish as $x \rightarrow \pm\infty$. \square

3.13 The proof above does not provide the best constant. Let $g = f'$ and let $g_+ = \min(0, g)$ and $g_- = \min(0, -g)$. Then $g = g_+ - g_-$. Assuming f has compact support, integrate g to show that $\int g_+ = \int g_- = \frac{1}{2} \int g_+ + g_- = \frac{1}{2} \int |f'|$. Show then that $\|f\|_{C^0} \leq \frac{1}{2} \|f'\|_{L^1}$.

3.14 Prove that $L_1^{k+1}(\mathbb{R}) \subset C_b^k(\mathbb{R})$ is an embedded subspace and that the first k derivatives of any $f \in L_1^{k+1}(\mathbb{R})$ vanish at $x = \pm\infty$.

Chapter 4

Fourier Transforms

Distributions are like functions but with mild singularities, sometimes singular enough that they can only be represented as “limiting behaviours” of functions. The Fourier transform of a function f is another function \hat{f} , which tells us how f is “built up” as a “sum” of sine and cosine waves of various frequencies.

Schwartz functions

A function f is *rapidly decreasing* if $x^a f$ is bounded for any a . A *Schwartz function* is a function f so that all of its derivatives $\partial^a f$ are rapidly decreasing. Let \mathcal{S} be the set of Schwartz functions. Clearly $C_c^\infty \subset \mathcal{S}$. The sum, difference and product of Schwartz functions is Schwartz. The product of a polynomial with a Schwartz function is Schwartz. If f is Schwartz, then $1 - e^f$, $\sin f$ and $\log(1 + |f|^2)$ are Schwartz, by the chain rule and l'Hôpital's rule.

4.1 Prove that $e^{-|x|^2} \in \mathcal{S}$.

4.2 Give an example of a function $f \in \mathcal{S}$ so that $e^{|x|^c} f$ is unbounded for any $c > 0$.

Let

$$\|f\|_{a,b} = \sup_x |x^a \partial^b f(x)|.$$

Say that a sequence of Schwartz functions f_1, f_2, \dots converges to a Schwartz function if and only if, for any a and b , $\|f - f_j\|_{a,b} \rightarrow 0$ as $j \rightarrow \infty$.

Lemma 4.1. *If $f, g \in \mathcal{S}$ and $f(0) = 1$ then*

$$f(\delta x)g(x) \rightarrow g(x) \text{ as } \delta \rightarrow 0.$$

Proof. Fix a positive integer N . Picking a large enough box (or ball) B so that, for any x outside B , as long as $|a| < N$ and $|b| < N$, we can ensure that all of the expressions $|x^a \partial^b g(x)|$ are as small as we like. If we now make δ small enough, then $f(\delta x) - 1$ is as small as we like inside the box B . Moreover, since every derivative of $f(\delta x) - 1$ has some factor of δ in it, we can ensure that these derivatives of order up to N are also as small as we like. Expanding out the derivatives $x^a \partial^b ((f(\delta x) - 1)g(x))$ using the chain rule, we get one factor or the other small throughout \mathbb{R}^n . \square

To ensure that sequences in \mathcal{S} converge as needed, we employ the metric

$$d(f, g) = \sum_{a,b} \frac{1}{2^{|a|+|b|}} \frac{\|f - g\|_{a,b}}{1 + \|f - g\|_{a,b}}.$$

Theorem 4.2. \mathcal{S} is a complete metric space.

Proof. A Cauchy sequence f_1, f_2, \dots converges uniformly on any compact set, with any number of derivatives, to some limit f . The f_j have rapidly decaying derivatives, so $x^a \partial^b f_j$ is bounded, and so $x^a \partial^b f$ is similarly bounded on each compact set. Moreover, $x^a \partial^b (f - f_j)$ gets small on that compact set, so for x in such a compact set

$$|x^a \partial^b (f - f_j)| = \lim_{k \rightarrow \infty} |x^a \partial^b (f_k - f_j)| \leq \lim_{k \rightarrow \infty} \|f_k - f_j\|_{a,b}.$$

But now make the compact set larger and larger, and you still get the same small bound of $\|f_k - f_j\|$,

$$\|f - f_j\|_{a,b} \leq \lim_{k \rightarrow \infty} \|f_k - f_j\|_{a,b}$$

and we can make this small by now making j get large. \square

Fourier transform

If $f \in L^1(\mathbb{R}^n)$, the *Fourier transform* of f is the function $\hat{f}: \mathbb{R}^n \rightarrow \mathbb{C}$ defined by

$$\begin{aligned} \hat{f}(\xi) &= \int_{\mathbb{R}^n} f(x) e^{-2\pi i \langle \xi, x \rangle}, \\ &= \left\langle f, e^{2\pi i \langle \xi, x \rangle} \right\rangle. \end{aligned}$$

The function $e^{2\pi i \langle \xi, x \rangle} = \cos(2\pi \langle \xi, x \rangle) + i \sin(2\pi \langle \xi, x \rangle)$ is a wave with ripples going up and down in the direction of ξ , of frequency $|\xi|$. Any inner product is a measure of how “correlated” or “sympathetic” two vectors are. So $\hat{f}(\xi)$ represents how much f is like a such a wave.

Lemma 4.3. *The Fourier transform of any integrable function is bounded; to be precise*

$$\|\hat{f}\|_{L^\infty} = \|f\|_{L^1}.$$

Proof. The Hölder inequality gives

$$|\hat{f}(\xi)| \leq \|f\|_{L^1} \|e^{-2\pi i \langle \xi, x \rangle}\|_{L^\infty} = \|f\|_{L^1}.$$

\square

Computing the Fourier transform of a Gaussian bell curve

4.3 Explain why

$$\int_{\mathbb{R}^2} e^{-x^2-y^2} = \left(\int_{\mathbb{R}} e^{-x^2} \right)^2.$$

Use polar coordinates to compute the left hand side. Use this to compute $\int_{\mathbb{R}} e^{-x^2} dx$. Use this to compute

$$\int_{\mathbb{R}} e^{-ax^2}.$$

Use this to compute

$$\int_{\mathbb{R}^n} e^{-\sum a_j x_j^2}.$$

Suppose that A is a positive definite symmetric matrix; orthogonally diagonalize to compute

$$\int_{\mathbb{R}^n} e^{-\langle Ax, x \rangle}.$$

To compute the Fourier transform of the function $f(x) = e^{-x^2}$, $f: \mathbb{R} \rightarrow \mathbb{R}$, write it as

$$\begin{aligned} \hat{f}(\xi) &= \int f(x) e^{-2\pi i \xi x}, \\ &= \int e^{-x^2 - 2\pi i \xi x}, \\ &= \int e^{-(x + \pi i \xi)^2 - \pi^2 \xi^2}, \\ &= e^{-\pi^2 \xi^2} \int e^{-(x + \pi i \xi)^2}. \end{aligned}$$

This integral can be written as an integral along a contour in the complex plane, say as

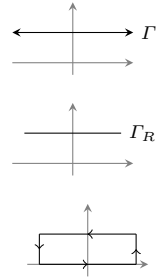
$$\int_{-\infty}^{\infty} e^{-(x + \pi i \xi)^2} dx = \int_{\Gamma} e^{-z^2} dz,$$

where Γ is the contour travelling along the line $z = x + \pi i \xi$, x going from $-\infty$ to ∞ .

We can approximate this contour by picking a large number, say R , and taking the same contour $z = x + \pi i \xi$ but only for $-R < x < R$.

Note that for large values of $|x|$, the function $e^{-z^2} = e^{-x^2 + 2ixy + y^2}$ decays faster than exponentially, so there is very little error in replacing $\int_{\Gamma} e^{-z^2} dz$ by $\int_{\Gamma_R} e^{-z^2} dz$. Consider the rectangle that has one side along Γ_R and another along the x -axis.

By Stokes's theorem, or the Cauchy integral theorem, because e^{-z^2} is holomorphic inside the rectangle, its integral around the boundary vanishes. The



left and right hand side of the rectangle sit in a region where, if we make R large, e^{-z^2} is smaller than an exponential decay in R , so we can make the rectangle very wide and find that the integral along the bottom becomes nearly the same as the integral along the top:

$$\lim_{R \rightarrow \infty} \int_{\Gamma_R} e^{-z^2} dz = \lim_{R \rightarrow \infty} \int_{-R}^R e^{-x^2} dx, \\ = \sqrt{\pi}.$$

Therefore $\hat{f}(\xi) = \sqrt{\pi} e^{-\pi^2 \xi^2}$: the Fourier transform of a Gaussian is another Gaussian.

4.4 Compute the Fourier transform \hat{f} of the Gaussian function $f: \mathbb{R}^n \rightarrow \mathbb{R}$,

$$f(x) = e^{-\langle Ax, x \rangle},$$

for any positive definite symmetric matrix A . You should find

$$\hat{f}(\xi) = \frac{\pi^{n/2}}{\sqrt{\det A}} e^{-\pi^2 \langle A^{-1} \xi, \xi \rangle}.$$

4.5 Use complex analysis to find Fourier transforms of some functions.

4.6 Compute

$f(x)$	$\hat{f}(\xi)$
$e^{- x }$	$\frac{2}{1+4\pi\xi^2}$
$\text{sgn}(x)e^{- x }$	$\frac{4\pi i \xi}{1+4\pi^2 \xi^2}$
$\max(0, 1-x)$	$\frac{e^{2\pi i \xi} - 1}{2\pi i \xi} \left(1 - \frac{1}{2\pi i \xi}\right)$

Properties of the Fourier transform on Schwartz functions

The Fourier transform $\mathcal{F}(f) = \hat{f}: L^1 \rightarrow L^\infty$ is a complex linear map. Differentiation under the integral sign shows that if $f \in \mathcal{S}$ then

$$\partial_j \hat{f}(\xi) = \int (-2\pi i x_j) f(x) e^{-2\pi i \langle \xi, x \rangle},$$

i.e. $\partial_j \mathcal{F}(f) = \mathcal{F}(-2\pi i x_j f)$. Differentiating several times,

$$\partial^a \mathcal{F}(f) = \mathcal{F}((-2\pi i x)^a f).$$

If p is any polynomial in n variables, say $p(x) = \sum c_a x^a$, then write $p(\partial)$ to mean $p(\partial) = \sum c_a \partial^a$. Then we have $p(\partial) \mathcal{F}(f) = \mathcal{F}(p(-2\pi i x) f)$. So \mathcal{F} turns

differentiation into polynomial multiplication. Similarly, if we differentiate,

$$\begin{aligned}\mathcal{F}(\partial_j f) &= \int \partial_j f(x) e^{-2\pi i \langle \xi, x \rangle}, \\ &= \int \partial_j \left(f(x) e^{-2\pi i \langle \xi, x \rangle} \right) - f(x) \partial_j e^{-2\pi i \langle \xi, x \rangle},\end{aligned}$$

to which we apply the fundamental theorem of calculus in one variable, since f vanishes at $x_j = \infty$:

$$\begin{aligned}&= - \int f(x) \partial_j e^{-2\pi i \langle \xi, x \rangle}, \\ &= - \int f(x) (-2\pi i \xi_j) e^{-2\pi i \langle \xi, x \rangle}, \\ &= 2\pi i \xi_j \mathcal{F}(f).\end{aligned}$$

Differentiating several times, $\mathcal{F}(p(\partial)f) = p(2\pi i \xi) \mathcal{F}(f)$ for any polynomial p . Roughly speaking, the Fourier transform interchanges differentiation in x with multiplication by a linear function in ξ and vice versa.

Lemma 4.4. *The Fourier transform $\mathcal{F}: \mathcal{S} \rightarrow \mathcal{S}$ is continuous.*

Proof. We saw that \mathcal{F} takes any integrable function to a bounded function. Consequently, the Fourier transform of a Schwartz function is Schwartz, $\mathcal{F}: \mathcal{S} \rightarrow \mathcal{S}$. Differences of Schwartz functions small in the norm $\|f - g\|_{a,b}$ are taken to differences of Schwartz functions small in the norm $\|\hat{f} - \hat{g}\|_{b,a}$. \square

4.7 Prove that when we translate or dilate

$$\begin{aligned}\mathcal{F}(f(x - x_0)) &= e^{-2\pi i \langle \xi, x_0 \rangle} \mathcal{F}(f(x)), \\ \mathcal{F}\left(e^{2\pi i \langle \xi_0, x \rangle} f(x)\right) &= \hat{f}(\xi - \xi_0), \\ \mathcal{F}(f(ax)) &= \frac{\hat{f}(\xi/a)}{|a|^n}.\end{aligned}$$

This last equation says that as f gets more “squished in”, \hat{f} gets more “spread out” and vice versa.

Lemma 4.5. *Gaussians are dense in the Schwartz functions.*

Proof. Take a Gaussian f on \mathbb{R}^n so that $\int f = 1$ and let $f_\varepsilon(x) = \varepsilon^{-n} f(x/\varepsilon)$. By theorem 2.7 on page 15, $f_\varepsilon * g \rightarrow g$ in L^∞ , i.e. uniformly. Since this holds for any Schwartz function, it also holds for the Schwartz function $x^a g$, for any a , and for the Schwartz function $x^a \partial^b g$. \square

The inverse Fourier transform

If $g \in L^1$, think of $g = g(\xi)$ as a function of ξ and let

$$\begin{aligned}\check{g}(x) &= \int g(\xi) e^{2\pi i \langle \xi, x \rangle}, \\ &= \left\langle g, e^{-2\pi i \langle \xi, x \rangle} \right\rangle, \\ &= \hat{g}(-x).\end{aligned}$$

Write the map $g \mapsto \check{g}$ as \mathcal{F}^* .

4.8 Prove that if $f(x) = e^{-\langle Ax, x \rangle}$ is a Gaussian, then $\mathcal{F}^* \mathcal{F} f = \mathcal{F} \mathcal{F}^* f = f$.

4.9 Suppose that $f \in \mathcal{S}$ satisfies $\mathcal{F}^* \mathcal{F} f = \mathcal{F} \mathcal{F}^* f = f$. Prove that $f(x - x_0)$ and $f(x/a)$ also satisfy this equation, for any $x_0 \in \mathbb{R}^n$ and $a \neq 0$.

Theorem 4.6. *If $f \in \mathcal{S}$ then $\mathcal{F}^* \mathcal{F} f = \mathcal{F} \mathcal{F}^* f = f$.*

Proof. It suffices to prove the result for translated and scaled Gaussians by lemma 4.5 on the previous page. You did this: problems 4.8 and 4.9. \square

Lemma 4.7. *If $f, g \in \mathcal{S}$ then*

$$\langle \hat{f}, g \rangle = \langle f, \check{g} \rangle.$$

Proof.

$$\begin{aligned}\langle \hat{f}, g \rangle &= \int \hat{f}(\xi) g(\xi) d\xi, \\ &= \int \left(\int f(x) e^{-2\pi i \langle \xi, x \rangle} dx \right) g(\xi) d\xi,\end{aligned}$$

apply Fubini's theorem,

$$\begin{aligned}&= \int f(x) \overline{\left(\int g(\xi) e^{2\pi i \langle \xi, x \rangle} d\xi \right)} dx \\ &= \int f(x) \overline{\check{g}(x)} dx, \\ &= \langle f, \check{g} \rangle.\end{aligned}$$

\square

4.10 If $f, g \in \mathcal{S}$, prove that $\mathcal{F}(f * g) = \hat{f} \hat{g}$.

4.11 Prove that multiplication $f \in \mathcal{S}, g \in \mathcal{S} \mapsto fg \in \mathcal{S}$ is continuous. Use this, and problem 4.10, to prove that convolution is continuous.

Fourier transforms of L^2 functions

Lemma 4.8. *If $f \in \mathcal{S}$ then $\|\mathcal{F}f\|_{L^2} = \|f\|_{L^2}$. In other words, $\mathcal{F}: \mathcal{S} \rightarrow \mathcal{S}$ is a unitary linear map.*

Proof.

$$\begin{aligned} \|\mathcal{F}f\|_{L^2}^2 &= \langle \mathcal{F}f, \mathcal{F}f \rangle, \\ &= \langle f, \mathcal{F}^* \mathcal{F}f \rangle, \\ &= \langle f, f \rangle, \\ &= \|f\|_{L^2}^2. \end{aligned}$$

□

Theorem 4.9 (Plancherel). *The Fourier transform admits a unique extension from $L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ to a unitary linear map $\mathcal{F}: L^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$.*

Proof. We would like to define \hat{f} for any $f \in L^2$ by writing f as a limit of a sequence of Schwartz functions, say f_1, f_2, \dots converging to f , and then letting $\hat{f} = \lim \hat{f}_j$. Since L^2 is a complete metric space, this limit will exist just when \hat{f}_j is a Cauchy sequence, which follows from f_j being Cauchy and \mathcal{F} being unitary. Suppose we pick a different sequence instead, say g_1, g_2, \dots ; by unitarity, $\hat{f}_j - \hat{g}_k \rightarrow 0$. So \hat{f} is well defined. Taking the limit $\hat{f}_j \rightarrow \hat{f}$, we easily see that $f \mapsto \hat{f}$ is complex linear. For any Schwartz function g ,

$$\langle \hat{f}, g \rangle = \lim \langle \hat{f}_j, g \rangle = \lim \langle f_j, \check{g} \rangle = \langle f, \check{g} \rangle.$$

Since $\mathcal{S} \subset L^2$ is dense, this same equation holds for $g \in L^2$. In particular, taking $g = f$, we see that $f \mapsto \hat{f}$ is continuous. □

Chapter 5

Distributions

Distributions are like functions but with mild singularities. They can be represented as “limiting behaviours” of sequences of smooth functions.

Definition

All of the measurements we make of the world are subject, at minute scales, to wave-like interferences. We can't measure with perfect precision. Instead of measuring the value $f(x)$ of a function, we measure smeared out “local averages”, i.e. integrals $\int f(x)g(x) dx$ where $g(x)$ might look like a little bump, so that the integral averages together values of f near the center of that bump. In this way, we can associate to each “bump function” $g(x)$ such an integral. If f is continuous, it is easy to prove that the map $g \mapsto \int f(x)g(x) dx$ determines f .

A *distribution* on an open set $U \subset \mathbb{R}^n$ is a linear map $\phi: C_c^\infty(U) \rightarrow \mathbb{C}$ which is continuous in the sense that if f_1, f_2, \dots converges to f in $C_c^\infty(U)$, and all of the functions f_1, f_2, \dots have support contained in a single compact set, then $\phi(f_1), \phi(f_2), \dots$ converges to $\phi(f)$. For example, any locally integrable function g defines a distribution $\phi(f) = \int f\bar{g}$; this is our most important example, and we will then denote ϕ as g and say that ϕ is a locally integrable function. Note that this would be silly if it were not true that any two locally integrable functions defining the same distribution must actually be equal. In imitation of this example, we usually write the expression $\phi(f)$ using a formal integral notation, as $\int f\bar{\phi}$ or as $\langle f, \phi \rangle$, as if ϕ were a locally integrable function. More exotically, any hypersurface S defines a distribution $\phi(f) = \int_S f$. The *Dirac delta function* is the distribution $\delta(f) = f(0)$. We can also define a distribution by $\phi(f) = \partial_i f(0)$. Weirdly, if $c \in \mathbb{C}$ and ϕ is a distribution, we write $c\phi$ to mean the distribution so that $\langle f, c\phi \rangle = \bar{c} \langle f, \phi \rangle$. This weird definition ensures that the distributions defined by locally integrable functions have “inner products” scaling correctly.

We proceed by following the analogy between locally integrable functions and distributions. If ϕ were a locally integrable function, and h were a C^∞ function, then $h\phi$ would be also locally integrable, and we would have $\langle f, h\phi \rangle = \langle f\bar{h}, \phi \rangle$, for all $f \in C_c^\infty$. Therefore we use this as a definition: if ϕ is a distribution and $h \in C^\infty$, we define $h\phi$ to be the distribution so that $\langle f, h\phi \rangle = \langle f\bar{h}, \phi \rangle$, for all $f \in C_c^\infty$. Similarly, if ϕ were a smooth function, integration by parts

would yield $0 = \int f \partial_i \bar{\phi} + \int \bar{\phi} \partial_i f$ since $f = 0$ outside a compact set inside our domain of integration, and therefore $\langle f, \partial_i \phi \rangle = -\langle \partial_i f, \phi \rangle$. Again we use this as a definition: if ϕ is a distribution, we denote by $\partial_i \phi$ the distribution defined by $\langle f, \partial_i \phi \rangle = -\langle \partial_i f, \phi \rangle$ for all $f \in C_c^\infty$. Similarly we define $\partial^a \phi$ to be the distribution so that $\langle f, \partial^a \phi \rangle = (-1)^{|a|} \langle \partial^a f, \phi \rangle$ for all $f \in C_c^\infty$. Any linear differential operator L on smooth functions has the form

$$Lu = \sum f_a \partial^a u.$$

If the coefficients f_a are smooth functions, we can define Lu for u a distribution by the same formula. We can define the adjoint L^* by

$$L^*u = \sum \bar{f}_a (-1)^{|a|} \partial^a u,$$

and we find $\langle Lf, g \rangle = \langle f, L^*g \rangle$ for $f \in C_c^\infty$ and g any distribution.

Distributions should be thought of as functions with mild singularities, not very nasty. For example, the function $1/x$ on \mathbb{R} is too singular to represent a distribution, i.e. $\langle f, 1/x \rangle$ is *not* defined for $f \in C_c^\infty$ unless $f(0) = 0$. It comes very close:

5.1 Prove that the expression

$$\langle f, \phi \rangle = \lim_{\varepsilon \rightarrow 0} \left(\int_{-\infty}^{-\varepsilon} \frac{f(x)}{x} dx + \int_{\varepsilon}^{\infty} \frac{f(x)}{x} dx \right)$$

is a distribution.

5.2 Prove that $|x|^\alpha$ is a distribution on \mathbb{R}^n as long as $\alpha + n > 0$ but not when $\alpha + n \leq 0$.

A distribution ϕ *vanishes* on an open set U if any test function f whose support lies in U has $\langle f, \phi \rangle = 0$. For example, δ vanishes on any open ball not containing the origin. The *support* of a distribution is the complement of the union of the open sets on which it vanishes.

5.3 Prove that the union of the open sets on which a distribution vanishes is an open set on which it vanishes.

Let $Rf(x) = \bar{f}(-x)$. By problem 2.3 on page 14, $\langle f * h, g \rangle = \langle f, Rh * g \rangle$ if f and g are test functions and $h \in L^1$. Therefore if ϕ is a distribution and h is a test function, we define $h * \phi = \phi * h$ to mean the distribution $\langle f, h * \phi \rangle = \langle Rh * f, \phi \rangle$.

5.4 Prove that $\partial^a (h * \phi) = (\partial^a h) * \phi = h * (\partial^a \phi)$ for any distribution ϕ and test function h .

Lemma 5.1. If ϕ is a distribution and f a test function, then $f * \phi$ is a smooth function given by

$$f * \phi(x) = \int f(x - y) \phi(y) dy,$$

where the right hand side is not actually an integral, but only a formal expression which means that we apply the distribution $\bar{\phi}(y)$ to the function $f(x - y)$.

Proof. As we vary x , $f(x - y)$ varies uniformly with any number of derivatives, and therefore our formal integral

$$I(x) = \int f(x - y)\phi(y) dy$$

(being in fact an application of a distribution to $f(x - y)$) varies continuously in x . For the moment, to simplify notation, pretend that \mathbb{R}^n is just \mathbb{R} . When we try to differentiate,

$$\frac{I(x + \Delta x) - I(x)}{\Delta x} = \int \left(\frac{f(x + \Delta x - y) - f(x - y)}{\Delta x} \right) \phi(y) dy$$

the difference quotient inside the integral converges uniformly on compact sets with any number of derivatives to $f'(x - y)$. So therefore $I(x)$ is differentiable. By induction, $I(x)$ is smooth. The same proof, with suitable notation, works in \mathbb{R}^n .

Pick any test function g and approximate $\langle g, I \rangle$ as a limit of Riemann sums: make a large box and cut it up into a grid of small boxes, say X_1, X_2, \dots, X_N , say with X_j having measure V_j , and take a point $x_j \in X_j$ in each box:

$$\begin{aligned} \langle g, I \rangle &\sim \sum g(x_j) \bar{I}(x_j) V_j = \sum g(x_j) V_j \int \bar{h}(x_j - y) \bar{\phi}(y) dy, \\ &= \int \sum g(x_j) V_j \int \bar{h}(x_j - y) \bar{\phi}(y) dy. \end{aligned}$$

But $\sum g(x_j) V_j \bar{h}(x_j - y) \rightarrow \int g(x) \bar{h}(x - y) dx = g * Rh(y)$ uniformly with any number of derivatives, because both g and h are test functions. So

$$\langle g, I \rangle = \langle g * Rh, \phi \rangle = \langle g, h * \phi \rangle.$$

□

A sequence ϕ_1, ϕ_2, \dots of distributions *converges* to a distribution ϕ if, for any test function f , $\langle f, \phi_j \rangle \rightarrow \langle f, \phi \rangle$.

Lemma 5.2. *Test functions are dense among distributions.*

Proof. Suppose that f and g are test functions on \mathbb{R}^n and that $\int f = 1$ and $g(0) = 0$. Let $f_\varepsilon(x) = \varepsilon^{-n} f(\varepsilon x)$, and let $\phi_\varepsilon = g(\varepsilon x) f_\varepsilon * \phi$. By lemma 5.1 on the facing page, ϕ_ε is a test function. We want to prove that $\phi_\varepsilon \rightarrow \phi$ as $\varepsilon \rightarrow 0$. For any test function h ,

$$\begin{aligned} \langle h, \phi_\varepsilon \rangle &= \langle h, g(\varepsilon x) f_\varepsilon * \phi \rangle, \\ &= \langle Rf_\varepsilon(g(\varepsilon x) * h), \phi \rangle. \end{aligned}$$

The proof of theorem 2.7 on page 15 is easily adjusted to prove that $Rf_\varepsilon(g(\varepsilon x) * h) \rightarrow h$ uniformly with any number of derivatives. □

Tempered distributions

By definition, distributions can be “integrated against” all smooth functions with compact support. We should expect that only some of the better behaved distributions can be “integrated against” a larger class of functions, like the Schwartz class. A *tempered distribution* is a continuous linear map $\phi: \mathcal{S} \rightarrow \mathbb{C}$. As before, we denote $\phi(f)$ as $\langle f, \phi \rangle$ or as $\int f \bar{\phi}$, and we define $c\phi$ by $\langle f, c\phi \rangle = \bar{c} \langle f, \phi \rangle$. To see if an operation ϕ defines a tempered distribution, we need to check continuity in all of the norms of \mathcal{S} , i.e. check that for any a, b there is some constant C so that

$$|\langle f, \phi \rangle| \leq C \sup_x |x^a \partial^b f|.$$

Each tempered distribution determines a distribution in the usual sense. By density of the test functions in the Schwartz functions, a distribution can only extend in at most one possible way from a continuous linear map on test functions to a continuous linear function on Schwartz functions. For example, e^x is a distribution, but *not* a tempered distribution. On the other hand, any locally integrable function growing more slowly than some polynomial is a tempered distribution. In particular, every L^p function is a tempered distribution for $1 \leq p \leq \infty$, and in particular Schwartz functions are tempered distributions. Every distribution with compact support is tempered. Define the product of a Schwartz function f and a tempered distribution ϕ by $\langle g, f\phi \rangle = \langle \bar{f}g, \phi \rangle$.

5.5 Is $e^{\ln(1+x^2)^2}$ a tempered distribution?

A sequence ϕ_1, ϕ_2, \dots *converges* to a tempered distribution ϕ if $\langle f, \phi_j \rangle \rightarrow \langle f, \phi \rangle$ for all $f \in \mathcal{S}$.

Lemma 5.3. *Every tempered distribution is the limit of a sequence of test functions.*

Proof. It is easy to adjust the proof of lemma 5.2 on the preceding page. \square

Clearly \mathcal{F}, ∂^a and multiplication by functions of at most polynomial growth all define continuous maps $\mathcal{S}' \rightarrow \mathcal{S}'$.

5.6 Compute $\hat{1}, \hat{\delta}, \hat{x^a}$.

5.7 If $f \in \mathcal{S}$, we have 3 definitions of $\mathcal{F}(f)$: directly as an integral, indirectly by treating f as a distribution, and indirectly by treating f as a tempered distribution. Prove that all 3 agree (in a suitable sense).

5.8 If $f \in \mathcal{S}$, we have 3 definitions of $\partial^a f$: directly as an integral, indirectly by treating f as a distribution, and indirectly by treating f as a tempered distribution. Prove that all 3 agree (in a suitable sense).

5.9 Prove that the convolution map $f \in \mathcal{S}, g \in \mathcal{S}' \mapsto f * g \in \mathcal{S}'$ is continuous.

By extension from \mathcal{S} , we find the obvious identities: for any $f \in \mathcal{S}$ and $g \in \mathcal{S}'$, $\langle f * g, h \rangle = \langle g, Rf * h \rangle$, $\partial^a f * g = (\partial^a f) * g = f * \partial^a g$ and $\mathcal{F}(f * g) = \hat{f}\hat{g}$.

Chapter 6

L^2 Theory of Derivatives

Will use functions with weak derivatives in L^2 in the study of differential equations.

Sobolev L^2 spaces and Sobolev embedding

Let L_k^2 be the set of all functions $f \in L^2$ so that f has weak derivatives (i.e. derivatives in the sense of distributions) $\partial^a f \in L^2$ for all $|a| \leq k$. On L_k^2 we define the inner product

$$\langle f, g \rangle_{L_k^2} = \sum_{|a| \leq k} \int \partial^a f \partial^a \bar{g}$$

and norm $\|f\|_{L_k^2} = \sqrt{\langle f, f \rangle_{L_k^2}}$.

Theorem 6.1. *With this inner product, L_k^2 is a Hilbert space, i.e. the norm is complete.*

Proof. Take a Cauchy sequence f_1, f_2, \dots in L_k^2 . If $|a| \leq k$, then $\partial^a f_1, \partial^a f_2, \dots$ converges in L^2 to some function, say f_a . Similarly, f_1, f_2, \dots converges in L^2 , say to f . We claim that f_a is a weak derivative $\partial^a f$ of f . Take any test function g :

$$\int (\partial^a f_j) g = (-1)^{|a|} \int f_j \partial^a g \rightarrow (-1)^{|a|} \int f \partial^a g,$$

but

$$\int (\partial^a f_j) g \rightarrow \int f_a g.$$

So $f_a = \partial^a f$ as distributions. □

Theorem 6.2. $C_c^\infty \subset \mathcal{S} \subset \dots \subset L_k^2 \subset \dots \subset L_1^2 \subset L_0^2 = L^2 \subset \mathcal{S}'$, with each space dense in all of the following spaces.

Proof. The inclusions are clear, and the density follows as long as we can prove that C_c^∞ is dense in L_k^2 , which we proved in theorem 3.4 on page 24. □

The *Japanese bracket* of a vector $x \in \mathbb{R}^n$ is $\langle x \rangle = \sqrt{1 + \|x\|^2}$. The importance of the Japanese bracket: it grows like $\|x\|$ as $\|x\| \rightarrow \infty$, but it is smooth everywhere.

Theorem 6.3. A tempered distribution $f \in \mathcal{S}'$ lies in L_k^2 if and only if $\langle \xi \rangle^k \hat{f} \in L^2$.

Proof. Suppose that $f \in L_k^2$. By Plancherel's theorem (theorem 4.9 on page 35), $\mathcal{F}(\partial^a f) \in L^2$ if $|a| \leq k$. But $\mathcal{F}(\partial^a f) = (2\pi i \xi)^a \hat{f}$, so $\hat{f}, |\xi|^k \hat{f} \in L^2$. On any ball B around the origin $\langle \xi \rangle$ is bounded and so $\langle \xi \rangle^k \hat{f} \in L^2(B)$. But if we make the ball big enough then, for ξ outside that ball, $\langle \xi \rangle^k \leq 2 \|\xi\|^k$, so $\langle \xi \rangle^k \hat{f} \in L^2$ outside the ball as well.

Conversely, suppose that $\langle \xi \rangle^k \hat{f} \in L^2$. Clearly $|\xi^a| \leq |\xi|^k \leq \langle \xi \rangle^k$, so $2\pi i \xi^a \hat{f} \in L^2$ for every a and therefore $f \in L_k^2$. \square

Generalize the Sobolev spaces: for any $s \in \mathbb{R}$, let $L_s^2(\mathbb{R}^n)$ be the set of all tempered distributions f so that $\langle \xi \rangle^s \hat{f} \in L^2(\mathbb{R}^n)$. We can identify any L_s^2 space with L^2 by

$$f \mapsto \mathcal{F}^{-1} \langle \xi \rangle^{-s} \mathcal{F} f.$$

So they are really all just L^2 in disguise. Clearly if $s \geq t$ then $L_s^2 \subset L_t^2$ is a dense embedded subspace. Each $f \in L_s^2$ is a tempered distribution, because $\int f \bar{g}$ is well defined for all $g \in L_{-s}^2$ and $\mathcal{S} \subset L_{-s}^2$.

6.1 Prove that $e^{-|x|} \in L_s^2(\mathbb{R})$ just when $s < \frac{3}{2}$.

Lemma 6.4. How big is the Japanese bracket? In \mathbb{R}^n , $\langle \xi \rangle^s \in L^2$ just when $s < -n/2$.

Proof. Integrate:

$$\begin{aligned} \|\langle \xi \rangle^s\|_{L^2}^2 &= \int \langle \xi \rangle^{2s}, \\ &= \int (1 + \|\xi\|^2)^s, \end{aligned}$$

polar coordinates: $\xi = ru$ where $r \geq 0$ and u is a unit vector

$$= \omega_{n-1} \int (1 + r^2)^s r^{n-1} dr$$

where ω_{n-1} is the hypersurface volume of the unit sphere in \mathbb{R}^n . The power of r is roughly $2s + n - 1$, so finite integral for $2s + n - 1 < -1$. \square

6.2 Prove that $\delta \in L_s^2(\mathbb{R}^n)$ just when $s < -\frac{n}{2}$. Express $\|\delta\|_{L_s^2}$ as an integral involving s . (The integral doesn't have an expression in elementary terms.)

Theorem 6.5 (Sobolev Embedding for L^2 Sobolev spaces). If $s - \frac{n}{2} > k$, or in other words if $\sigma L_s^2 > \sigma C^k$, then $L_s^2(\mathbb{R}^n) \subset C_b^k(\mathbb{R}^n)$.

Proof. For any test function f , if $|a| = k$,

$$\begin{aligned}
 |\partial^a f| &= \left| \int e^{2\pi i \langle \xi, x \rangle} (2\pi i \xi)^a \hat{f}(\xi) d\xi \right|, \\
 &\leq \int |(2\pi i \xi)^a \hat{f}(\xi)| d\xi, \\
 &= \int |(2\pi i \xi)^a \hat{f}(\xi)| \langle \xi \rangle^{s-k} \langle \xi \rangle^{k-s} d\xi, \\
 &\leq \left\| (2\pi i \xi)^a \langle \xi \rangle^{s-k} \hat{f} \right\|_{L^2} \left\| \langle \xi \rangle^{k-s} \right\|_{L^2}, \\
 &\leq C \|\partial^a f\|_{L^2_{s-k}}.
 \end{aligned}$$

□

Corollary 6.6. *A function is smooth with square integrable derivatives of all orders just when it belongs to all Sobolev L^2 spaces.*

Theorem 6.7. *A function $f \in L^2_s(\mathbb{R}^n)$ lies in $L^2_{s+1}(\mathbb{R}^n)$ if and only if the difference quotient*

$$\frac{f(x + hv) - f(x)}{h}$$

is bounded in L^2 as $h \rightarrow 0$ for any constant vector $v \in \mathbb{R}^n$ and $h \in \mathbb{R}$.

Proof. If $f \in L^2_{k+1}(\mathbb{R}^n)$ then clearly the difference quotient converges to $\langle df, v \rangle$. Suppose that the difference quotient is bounded in L^2 . Then its Fourier transform is also bounded in L^2 by the Plancherel theorem. Compute that

$$\begin{aligned}
 \mathcal{F}\left(\frac{f(x + hv) - f(x)}{h}\right) &= \frac{e^{2\pi i \langle \xi, hv \rangle} - 1}{h} \hat{f}, \\
 &= \underbrace{\frac{e^{2\pi i \langle \xi, hv \rangle} - 1}{2\pi i \langle \xi, hv \rangle}}_{\rightarrow 1 \text{ as } h \rightarrow 0} \underbrace{2\pi i \langle \xi, v \rangle \hat{f}}_{\mathcal{F}\langle df, v \rangle},
 \end{aligned}$$

and the bounded factor goes to 1 pointwise. Because this is bounded in L^2 as $h \rightarrow 0$, we can apply the dominated convergence theorem:

$$\rightarrow \mathcal{F}\langle df, v \rangle.$$

Therefore $\mathcal{F}\langle df, v \rangle \in L^2$, and so by Plancherel's theorem again $\langle df, v \rangle \in L^2$. □

Trace

If $X \subset \mathbb{R}^n$ is a subset and f is a continuous function defined near X , it is traditional to write the restriction $f|_X$ as $\text{tr}_X(f)$. If X has measure zero, and f is only defined up to a set of measure zero, then the trace is not defined.

If $X \subset \mathbb{R}^n$ is a linear subspace (or an affine subspace, i.e. a translate of a linear subspace) of dimension k , we can rotate and translate X into $\mathbb{R}^k \times \{0\}$, and define the Sobolev and Hölder spaces of X as those of \mathbb{R}^k . We can assume that $X = \mathbb{R}^k \times \{0\}$ and write each point of \mathbb{R}^n as (x, y) . Similarly we can write the coordinates for the Fourier transform as (ξ, η) . To understand traces on Sobolev spaces, we relate Japanese brackets of linear subspaces to those of ambient spaces.

Lemma 6.8. *For any fixed $\xi \in \mathbb{R}^k$ and variable $\eta \in \mathbb{R}^{n-k}$ and $s > (n-k)/2$, there is a constant C so that for all ξ ,*

$$\int \langle \xi, \eta \rangle^{-2s} d\eta = C \langle \xi \rangle^{-2s+n-k}.$$

Proof.

$$\int \langle \xi, \eta \rangle^{-2s} d\eta = \int \left(1 + \|\xi\|^2 + \|\eta\|^2\right)^{-s} d\eta.$$

Let $a = \langle \xi \rangle = \sqrt{1 + \|\xi\|^2}$ and $r = \|\eta\|$, and use “polar coordinates” in η , taking ω_{n-k-1} to be the hypersurface area of the unit sphere in \mathbb{R}^{n-k} :

$$\int \langle \xi, \eta \rangle^{-2s} d\eta = \omega_{n-k-1} \int (a^2 + r^2)^{-s} r^{n-k-1} dr,$$

and now let $u = r/a$

$$\begin{aligned} &= \omega_{n-k-1} \int a^{-2s} (1 + u^2)^{-s} a^{n-k-1} u^{n-k-1} a du, \\ &= \omega_{n-k-1} a^{-2s+n-k} \int (1 + u^2)^{-s} u^{n-k-1} du. \end{aligned}$$

If $-2s + n - k - 1 < -1$ this integral converges. Plug in $a = \langle \xi \rangle$. \square

Theorem 6.9. *Suppose that $A \subset \mathbb{R}^n$ is an affine subspace of dimension k . Consider the trace map $\text{tr}_A: f \in \mathcal{S}(\mathbb{R}^n) \mapsto f|_A \in \mathcal{S}(A)$. If $\sigma L_s^2(\mathbb{R}^n) \geq \sigma L_t^2(A)$, in other words $s - \frac{n}{2} \geq t - \frac{k}{2}$, then tr_A extends to a unique continuous linear map $\text{tr}_A: L_s^2(\mathbb{R}^n) \rightarrow L_t^2(A)$. If $\sigma L_s^2(\mathbb{R}^n) = \sigma L_t^2(A)$, then this linear map is surjective.*

Proof. Take $f \in \mathcal{S}(\mathbb{R}^n)$ and let $g = \text{tr}_X(f)$. We only need to find a constant C so that $\|g\|_{L_t^2} \leq C \|f\|_{L_s^2}$. We can assume that $X = \mathbb{R}^k \times \{0\}$ and write each point of \mathbb{R}^n as (x, y) . Similarly we can write the coordinates for the Fourier transform as (ξ, η) . Equivalently we only need to ensure that

$$\left\| \langle \xi \rangle^t \hat{g} \right\|_{L^2} \leq C \left\| \langle \xi, \eta \rangle^s \hat{f} \right\|_{L^2}.$$

Let's relate \hat{g} to \hat{f} .

$$\begin{aligned} g(x) &= f(x, 0), \\ &= \int e^{2\pi i \langle (\xi, \eta), (x, 0) \rangle} \hat{f}(\xi, \eta) d\xi d\eta, \\ &= \int e^{2\pi i \langle \xi, x \rangle} \hat{f}(\xi, \eta) d\xi d\eta. \end{aligned}$$

Therefore

$$\begin{aligned} \hat{g}(\xi) &= \int \hat{f}(\xi, \eta) d\eta, \\ &= \int \hat{f}(\xi, \eta) \langle \xi, \eta \rangle^s \langle \xi, \eta \rangle^{-s} d\eta. \end{aligned}$$

By Hölder's inequality,

$$|\hat{g}(\xi)| \leq \left\| \hat{f}(\xi, \eta) \langle \xi, \eta \rangle^s \right\|_{L^2(\eta)} \left\| \langle \xi, \eta \rangle^{-s} \right\|_{L^2(\eta)},$$

and we apply the previous lemma

$$\leq \sqrt{C} \langle \xi \rangle^{-s+(n-k)/2} \left\| \hat{f}(\xi, \eta) \langle \xi, \eta \rangle^s \right\|_{L^2(\eta)}.$$

Therefore the Sobolev norm of g is given by

$$\begin{aligned} \|g\|_{L^2_t}^2 &= \int |\hat{g}(\xi)|^2 \langle \xi \rangle^{2t} d\xi, \\ &\leq C \int \langle \xi \rangle^{2t-2s+n-k} \left\| \hat{f}(\xi, \eta) \langle \xi, \eta \rangle^s \right\|_{L^2(\eta)}^2 d\xi, \\ &= C \int \langle \xi \rangle^{2t-2s+n-k} \left| \hat{f}(\xi, \eta) \right|^2 \langle \xi, \eta \rangle^{2s} d\eta d\xi, \end{aligned}$$

but $\langle \xi \rangle^{2t-2s+n-k} \leq 1$ so

$$\begin{aligned} &\leq C \int \left| \hat{f}(\xi, \eta) \right|^2 \langle \xi, \eta \rangle^{2s} d\eta d\xi, \\ &= C \|f\|_{L^2_s}^2. \end{aligned}$$

To see that tr_X is onto, it is sufficient to prove the result for $s - \frac{n}{2} = t - \frac{k}{2}$, given $g(x)$, let

$$\hat{f}(\xi, \eta) = \frac{1}{C^2} \hat{g}(\xi) \langle \xi \rangle^{t-k/2} \langle \xi, \eta \rangle^{n/2-s}$$

and let $f = \mathcal{F}^{-1} \hat{f}$. □

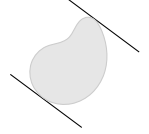
Chapter 7

The Direct Method of the Calculus of Variations

We will try to find the functions which have least energy in an appropriate sense.

Bounding the value by the derivative

Pick a bounded set $A \subset \mathbb{R}^n$ and a unit vector $u \in \mathbb{R}^n$. The *width* of A in the direction u is the smallest distance between two parallel hyperplanes normal to u ; if A is compact then the width depends continuously on u .



The *minimum width* of A is the minimum of the width in any direction.

Lemma 7.1. *If U is a compact domain with smooth boundary and minimum width m , then every $u \in C^1(U)$ with $u|_{\partial U} = 0$ satisfies*

$$\|u\|_{L^2} \leq m \|du\|_{L^2}.$$

Proof. Let $a = m/2$ and rotate and translate U to arrange that $-a \leq x_1 \leq a$ at every point of U . Extend u to vanish outside U . Write each point $x \in \mathbb{R}^n$ as $x = (s, t)$ with $s \in \mathbb{R}, t \in \mathbb{R}^{n-1}$. Then at each point (s_0, t_0) with $s_0 \leq 0$,

$$\begin{aligned} u(s_0, t_0)^2 &= \int_{-a}^{s_0} \partial_s (u(s, t_0)^2) ds, \\ &= 2 \int_{-a}^{s_0} u(s, t_0) \partial_s u(s, t_0) ds, \\ &\leq 2 \left(\int_{-a}^0 u(s, t_0)^2 ds \int_{-a}^0 (\partial_s u(s, t_0))^2 ds \right)^{1/2}. \end{aligned}$$

The right hand side is a constant in s . Integrate in s :

$$\int_{-a}^0 u(s, t)^2 ds \leq 2a \left(\int_{-a}^0 u(s, t_0)^2 ds \int_{-a}^0 (\partial_s u(s, t_0))^2 ds \right)^{1/2}.$$

Square both sides:

$$\int_{-a}^0 u(s, t)^2 ds \leq (2a)^2 \int_{-a}^0 (\partial_s u(s, t))^2 ds.$$

Integrate in t :

$$\int_{\mathbb{R}^{n-1}} \int_{-a}^0 u^2(s, t) ds dt \leq (2a)^2 \int_{\mathbb{R}^{n-1}} \int_{-a}^0 (\partial_1 u)^2(s, t) ds dt,$$

and adding in the other half, where $0 \leq s \leq a$,

$$\int u^2 \leq (2a)^2 \int (\partial_1 u)^2 \leq (2a)^2 \int \|du\|^2 = m^2 \int \|du\|^2.$$

□

The variational problem

Pick a compact domain $U \subset \mathbb{R}^n$ with smooth boundary. To each smooth real-valued function $u \in C^\infty(U)$, associate the number

$$S[u] = \int_U \left(\frac{1}{2} \|du\|^2 + f(x)u(x) \right)$$

where $f: U \rightarrow \mathbb{R}$ is a smooth function. Call this the *action* of a function u . Among all functions u which vanish on ∂U , let us try to find one which makes the action as small as possible.

Changing the boundary values

Consider a slightly more general problem. Fix a smooth function $h \in C^\infty(\partial U)$. Among all functions u which equal h on ∂U , let us try to find one which makes the action as small as possible. Pick one smooth function u_0 which equals h on ∂U . Then write every other such function as $u = u_0 + v$. So the functions v are just those which vanish on ∂U .

7.1 Use the divergence theorem to compute that $S[u_0 + v] = S[u_0] + T[v]$ where

$$T[v] = \int_U \left(\frac{1}{2} \|dv\|^2 + k(x)v(x) \right)$$

for some function $k(x)$.

But then S is minimal at some $u = u_0 + v$ just when T is minimal at v . So if we can solve our origin problem (with $u = 0$ on ∂U), for any action functional S , then we can solve this more general problem.

7.2 If we add a term linear in the first derivatives, say let

$$S[u] = \int_U \left(\frac{1}{2} \|du\|^2 + f(x)u(x) + \sum_i h_i(x) \partial_i u \right)$$

where $h_1(x), h_2(x), \dots, h_n(x)$ are smooth functions, integrate by parts to show that we can rearrange the action to have no such terms.

Sobolev spaces vanishing on the boundary

Let $\mathring{L}_k^2(U)$ be the closure in $L_k^2(U)$ of the smooth functions on U which vanish near ∂U . By the Sobolev embedding theorem, $L_1^2(U) \subset C^0(U)$, so the functions in $\mathring{L}_1^2(U)$ are continuous and vanish on ∂U .

Bounding the action

Lemma 7.2. *Among all real-valued functions $u \in \mathring{L}_1^2(U)$, the values of $S[u]$ are bounded from below.*

Proof.

$$\begin{aligned} 0 &\leq \frac{1}{2} \int_U \|du\|^2, \\ &\leq S[u] - \int_U f(x)u(x) \end{aligned}$$

So if $S[u]$ gets arbitrarily large negative, then, to compensate

$$\int_U f(x)u(x) dx$$

must also get arbitrarily large negative. By Hölder's inequality, $\|f\|_{L^2} \|u\|_{L^2}$ must get arbitrarily large positive. Applying the inequality

$$ab \leq \frac{a^2 + b^2}{2},$$

we find, for any $\varepsilon > 0$,

$$\|f\|_{L^2} \|u\|_{L^2} = \left(\frac{\|f\|_{L^2}}{\sqrt{\varepsilon}} \right) (\sqrt{\varepsilon} \|u\|_{L^2}) \leq \frac{1}{2\varepsilon} \|f\|_{L^2}^2 + \frac{\varepsilon}{2} \|u\|_{L^2}^2.$$

If U has minimum width m then

$$\|u\|_{L^2} \leq m \|du\|_{L^2},$$

so

$$-\frac{\varepsilon}{2} \|u\|_{L^2}^2 \geq -\frac{\varepsilon m}{2} \|du\|_{L^2}^2.$$

Therefore

$$\begin{aligned}
S[u] &\geq \frac{1}{2} \|du\|_{L^2}^2 - \int_U f(x)u(x), \\
&\geq \frac{1}{2} \|du\|_{L^2}^2 - \frac{1}{2\varepsilon} \|f\|_{L^2}^2 - \frac{\varepsilon}{2} \|u\|_{L^2}^2, \\
&\geq \frac{1}{2} \|du\|_{L^2}^2 - \frac{1}{2\varepsilon} \|f\|_{L^2}^2 - \frac{\varepsilon m}{2} \|du\|_{L^2}^2, \\
&\geq \frac{(1 - \varepsilon m)}{2} \|du\|_{L^2}^2 - \frac{1}{2\varepsilon} \|f\|_{L^2}^2, \\
&\geq -\frac{1}{2\varepsilon} \|f\|_{L^2}^2,
\end{aligned}$$

a bound independent of the choice of u , as long as we pick ε so that $\varepsilon m < 1$. \square

Lemma 7.3. *Suppose that U is a compact domain with smooth boundary. For functions in $\mathring{L}_1^2(U)$, any bound on the action $S[u]$ imposes a bound on the Sobolev norm $\|u\|_{L_1^2}$.*

Proof. As in the proof of lemma 7.2 on the previous page,

$$\frac{(1 - m\varepsilon)}{2} \|du\|_{L^2}^2 \leq S[u] + \frac{1}{2\varepsilon} \|f\|_{L^2}^2,$$

so $\|du\|_{L^2}$ is bounded. By lemma 7.1 on page 47, $\|u\|_{L^2}$ is bounded, and so $\|u\|_{L_1^2}^2 = \|u\|_{L^2}^2 + \|du\|_{L^2}^2$ is bounded. \square

Lemma 7.4. *If $u_j \rightarrow u$ weakly in L_1^2 , then $\|u\|_{L_1^2} \leq \liminf \|u_j\|_{L_1^2}$.*

Proof.

$$\begin{aligned}
\|u\|_{L_1^2}^2 &= \langle u, u \rangle_{L_1^2}, \\
&= \lim_j \langle u_j, u \rangle_{L_1^2}, \\
&\leq \liminf_j \left| \langle u_j, u \rangle_{L_1^2} \right|, \\
&\leq \liminf_j \|u_j\|_{L_1^2} \|u\|_{L_1^2},
\end{aligned}$$

and we divide both sides by $\|u\|_{L_1^2}$. \square

Lemma 7.5. *Suppose that U is a bounded domain with smooth boundary. There is a function $u \in \mathring{L}_1^2(U)$ and vanishing on ∂U so that $S[u] = \inf_v S[v]$, infimum among all functions in \mathring{L}_1^2 .*

Proof. Take a sequence of functions $u_j \in C^\infty(U)$ with $u_j|_{\partial U} = 0$, so that the values $S[u_j]$ of the action approach the infimum value. By the previous lemma, since the functions u_j have bounded action, they have bounded Sobolev norm

$\|u_j\|_{L^2_1}$, so the sequence u_j is bounded in $L^2_1(U)$. In problem 3.5 on page 23, we saw that u_j has a weakly convergent subsequence in L^2_1 ; replace u_j by that subsequence.

By the Kondrashov–Rellich theorem (theorem 3.7 on page 27) there is a subsequence of the u_j that converges in L^p if

$$1 \leq p < \begin{cases} \infty, & \text{if } n = 2, \\ \frac{2n}{n-2} & \text{if } n \geq 3; \end{cases}$$

for any finite set of values of p we can replace the u_j by that subsequence. Again by the Kondrashov–Rellich theorem, there is a subsequence of the u_j that converges in $C^{0,\alpha}$ if $0 < \alpha < 1 - \frac{n}{2}$; again replace these u_j by that subsequence for any finite set of values of α . So now u_j converges in L^p for small enough p and in $C^{0,\alpha}$ for small enough α and weakly in L^2_1 . These various function spaces are all contained in L^2 , and there the various limits must all agree as L^2 functions, so as distributions. So $u_j \rightarrow u$ in L^p and $C^{0,\alpha}$ and L^2_1 for the appropriate range of p and α .

We need to prove that $S[u] = \lim_j S[u_j]$. This is not obvious, because S is perhaps not a continuous function on L^2_1 . But

$$\begin{aligned} S[u_j] &= \int \frac{1}{2} \|du_j\|^2 + \int f u_j, \\ &= \frac{1}{2} \int \|du_j - du\|^2 + \int \langle du_j, du \rangle - \frac{1}{2} \int \|du\|^2 + \int f u_j, \end{aligned}$$

The first term is nonnegative and it is the nonlinear part in u_j : drop it to get a smaller value:

$$\geq \int \langle du_j, du \rangle - \frac{1}{2} \int \|du\|^2 + \int f u_j,$$

The second term is constant in j , while the first and third are applying continuous linear functions to u_j , so we can take the limit:

$$\begin{aligned} &\rightarrow \int \langle du, du \rangle - \frac{1}{2} \int \|du\|^2 + \int f u, \\ &= S[u]. \end{aligned}$$

□

Lemma 7.6. For any $u, v \in L^2_1$,

$$S\left[\frac{u+v}{2}\right] \leq \frac{S[u] + S[v]}{2}.$$

Proof. For any L^2 functions u and v , by Hölder's inequality

$$\int \langle du, dv \rangle \leq \left(\int \|du\|^2 \right)^{1/2} \left(\int \|dv\|^2 \right)^{1/2}$$

with equality just when $dv = c du$ or $du = c dv$ for some constant $c \geq 0$. By the arithmetic geometric mean inequality ($ab \leq (1/2)(a + b)$, equality just when $a = b$),

$$\int \langle du, dv \rangle \leq \frac{1}{2} \left(\int \|du\|^2 + \int \|dv\|^2 \right)$$

with equality just when $du = dv$, which implies $u = v$ since u and v agree on ∂U . Therefore

$$\begin{aligned} S \left[\frac{u+v}{2} \right] &= \int \frac{1}{2} \left\| \frac{du+dv}{2} \right\|^2 + \frac{1}{2} \int fu + \frac{1}{2} \int fv, \\ &= \frac{1}{8} \int \|du\|^2 + \frac{1}{4} \int \langle du, dv \rangle + \frac{1}{8} \int \|dv\|^2 + \frac{1}{2} \int fu + \frac{1}{2} \int fv, \end{aligned}$$

to which we apply the previous estimate:

$$\begin{aligned} &\leq \frac{1}{8} \int \|du\|^2 + \frac{1}{8} \int \|du\|^2 + \frac{1}{8} \int \|dv\|^2 + \frac{1}{8} \int \|dv\|^2 + \frac{1}{2} \int fu + \frac{1}{2} \int fv, \\ &= \frac{S[u] + S[v]}{2}. \end{aligned}$$

□

Lemma 7.7. *The function u which minimizes the action S among all functions in \mathring{L}_1^2 is unique.*

Proof. Suppose that u and v are minimizers of action. By the previous lemma, $(1/2)(u + v)$ has action no larger, so must have equal action. Reversing the steps in the proof of that lemma, we must have equality everywhere. Equality in the Hölder inequality forces $du = c dv$ or $dv = c du$ for some $c > 0$. Equality in the arithmetic geometric mean inequality forces

$$\int \|du\|^2 = \int \|dv\|^2,$$

which forces $c = 1$.

□

Lemma 7.8. *The following are equivalent for a function $u \in \mathring{L}_1^2(U)$:*

- a. u is a weak solution of $\Delta u = f$,
- b. $0 = \int \langle du, dv \rangle + fv$ for any $v \in C_c^\infty$,
- c. $0 = \int \langle du, dv \rangle + fv$ for any $v \in \mathring{L}_1^2(U)$,
- d. u is a critical point of the action on $\mathring{L}_1^2(U)$,
- e. u is the minimizer of the action on $\mathring{L}_1^2(U)$.

Proof. The definition of Δu is that $\int (\Delta u)v = -\int \langle du, dv \rangle$ for any $v \in C_c^\infty$. So $\langle f - \Delta u, v \rangle = \int \langle du, dv \rangle + \int f v$. Hence (1) is equivalent to (2). Clearly (3) implies (2) because $C_c^\infty(U) \subset \mathring{L}_1^2(U)$. But (2) implies (3) by density of $C_c^\infty(U)$ in $\mathring{L}_1^2(U)$.

Suppose that u is a minimizer of the action. For any $v \in \mathring{L}_1^2(U)$, $S[u + tv]$ increases or stays constant as we vary t away from $t = 0$, so

$$\begin{aligned} 0 &\leq S[u + tv] - S[u], \\ &= \int \frac{1}{2} \|du + t dv\|^2 + \int f(u + tv) - \int \frac{1}{2} \|du\|^2 + \int f u, \\ &= t^2 \int \frac{1}{2} \|dv\|^2 + t \int \langle du, dv \rangle + t \int f v \end{aligned}$$

This holds for t positive and negative; for $t > 0$ divide by t and then send $t \rightarrow 0$ to get

$$0 \leq \int \langle du, dv \rangle + \int f v.$$

For $t < 0$, do the same to get

$$0 \geq \int \langle du, dv \rangle + \int f v.$$

So

$$0 = \int \langle du, dv \rangle + \int f v.$$

The same derivation shows that S is critical at u just when (3) is satisfied, in the sense that

$$0 = \lim_{t \rightarrow 0} \frac{S[u + tv] - S[u]}{t}.$$

If u is a minimizer, then u is critical because S is differentiable as above. If there are two critical functions u and w for S , then

$$0 = \int \langle du, dv \rangle + \int f v = \int \langle dw, dv \rangle + \int f v$$

for any $v \in \mathring{L}_1^2(U)$, so

$$0 = \int \langle du - dw, dv \rangle,$$

and if we let $v = u - w$,

$$0 = \int \|du - dw\|^2,$$

so that $du = dw$. By lemma 7.1 on page 47, $u = w$. □

Consequently, there is a unique solution u in $\mathring{L}_1^2(U)$ to $\Delta u = f$.

Chapter 8

Linear Elliptic Second Order Partial Differential Equations

We describe the basic intuitions of linear second order partial differential equations.

Physical intuition

8.1 Recall that a symmetric matrix A is *positive definite* if all of the eigenvalues of A are positive. Recall also that every symmetric matrix A has an orthonormal basis of eigenvectors. Prove that any symmetric matrix A is positive definite just when $\langle Ax, x \rangle > 0$ for any vector $x \neq 0$. Use this to prove that if A and B are both positive definite, then $\text{tr}(AB) > 0$.

Suppose that u is the temperature at location x at time t . Over time, the temperature changes, usually according to a partial differential equation looking something like

$$\partial_t u = \sum_{ij} a_{ij}(x) \partial_{ij} u + \sum_i X_i(x) \partial_i u + f(x)u + g(x).$$

What do the various terms represent? First, $\partial_t u$ is the rate at which u changes over time. For simplicity, the right hand side coefficients only depend on x because we imagine that they represent physical phenomena that don't change over time. Start with the last term: g . Imagine this was the only term. If $g > 0$, $\partial_t u = g > 0$, so u goes up. So g is a heater, like a stove or a radiator. If $g < 0$, g is a refrigerator or a block of ice. Next, imagine that $f(x)u$ was the only term: $\partial_t u = f(x)u$. Then the solution is $u(t, x) = e^{tf(x)}u(0, x)$, exponential growth or decay at rate $f(x)$. So $f(x)u$ is a term that “snowballs” the heat where $f > 0$ and decays heat away where $f < 0$. (This sort of term seems less well motivated physically.) Next, imagine that $\sum_i X_i(x) \partial_i u$ was the only term:

$$\partial_t u = \sum_i X_i(x) \partial_i u.$$

Consider the *vector field*

$$X(x) = \begin{pmatrix} X_1(x) \\ X_2(x) \\ \vdots \\ X_n(x) \end{pmatrix}.$$

The *flow lines* of this vector field are the curves $x(t)$ in space for which $x'(t) = X(x(t))$, i.e. the “particle” $x(t)$ moves so that its velocity at each moment in time agrees with X at its location. By the Picard existence and uniqueness theorem for ordinary differential equations, if X is C^1 then there is a unique C^2 flow line through each point. By the chain rule, u satisfies $\partial_t u = \sum_i X_i(x) \partial_i u$ just when u is constant along the flow lines. So this term represents carrying the heat along the flow of X : the vector field flows the heated molecule along. Finally, the first and worst term: suppose that $\partial_t u = \sum_{ij} a_{ij} \partial_{ij} u$. It turns out that this type of term represents the “diffusion” of the heat, as it spreads out in space, carried by the random motions of molecules. The eigenvalues of the matrix $A = (a_{ij})$ will turn out to tell us how rapidly the heat wants to spread out in each direction. To make this work, we need to assume that $A(x)$ is a positive definite matrix at each point x . At a point where u has a peak, any peak is a critical point so $\partial_i u = 0$ for all i , and the matrix $\partial^2 u = (\partial_{ij} u)$ of second derivatives will have negative eigenvalues. So then $\partial_t u = \sum_{ij} a_{ij} \partial_{ij} u = \text{tr } A \partial^2 u$ is the trace of a product of a positive definite matrix and a negative definite one, so a negative number. In other words, near a peak of u , u goes down over time: the hottest spots cool. Changing sign in the argument, putting it upside down, the coolest spots heat up: temperature wants to equilibrate. Suppose that $A = (a_{ij})$ is constant, for simplicity. Let Y be the vector field $Y_i(x) = -\sum_j a_{ij} \partial_j u = -A du$. For any bounded open set U with C^1 boundary,

$$\begin{aligned} \frac{d}{dt} \int_U u &= \int_U \partial_t u, \\ &= \int_U \sum_{ij} a_{ij} \partial_{ij} u, \\ &= \int_U \sum_i \partial_i Y, \end{aligned}$$

and if n is the unit normal vector to ∂U then by Stokes’s theorem

$$= \int_{\partial U} \langle Y, n \rangle.$$

Roughly: Y pushes u out of U . Since A has positive eigenvalues, then for any eigenvector v of A , we see that $-Av$ points the opposite direction to v . So roughly, the vector Y points in almost the opposite direction to du . Roughly, this says that the heat flows in almost the opposite direction to du . Picture the graph of u as a landscape, and imagine standing on it. The direction of du is uphill. If there is a hill with a steep slope to your left, then u is flowing to your right, making that steep hill of u get smaller, so flowing from larger to smaller. This is the nature of diffusion: heat spreads out, heating up cold things, with heat drawn away from hot things.

If we wait long enough, perhaps our heat will eventually settle into an equilibrium, with a fixed temperature function. Then $u = u(x)$ doesn’t change

in time anymore, and our equation is now

$$0 = \sum_{ij} a_{ij}(x) \partial_{ij} u + \sum_i X_i(x) \partial_i u + f(x)u + g(x).$$

This is the equation we will study, as a first step in developing a theory of partial differential equations for use in mathematical physics.

Linearization

Given any sufficiently smooth nonlinear partial differential equation $P[u] = 0$, we can always approximate it with a linear equation as follows. Take any function v , and expand out

$$P[u + \varepsilon v] = P[u] + \varepsilon P'[u]v + o(\varepsilon).$$

For example, if

$$P[u] = \partial_{xx} u + u \partial_x u + u^4,$$

then

$$P'[u]v = \partial_{xx} v + u \partial_x v + v \partial_x u + 4u^3 v.$$

Note that this differential operator is linear in v , but depends on the choice of solution u to the nonlinear equation.

8.2 Linearize the equation $\Delta u = |du|^2$ around the solution $u = 1$.

8.3 Linearize the Liouville equation $\Delta \log u = -u^2$ around the solution $u = \frac{1}{2} \log \left(\frac{4}{(1-x^2-y^2)^2} \right)$.

Characteristics

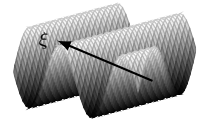
In Fourier transforms, we are always running into factors of $2\pi i$. It is convenient to write out all linear differential operators in terms of the operation $D_j = \partial_j / 2\pi i$ and $D^a = \partial^a / (2\pi i)^{|a|}$. For any constant coefficient linear differential operator $Q(D) = \sum_a c_a D^a$ in this D notation

$$\mathcal{F}(Q(D)f) = Q(\xi)\hat{f}.$$

A wave looks like the real or imaginary part of $e^{2\pi i \langle \xi, x \rangle}$, as a function of x for some fixed vector ξ , called the *momentum* of the wave. The momentum points in the direction that the wave ripples up and down, perpendicular to the directions where the wave has constant height.

Pick a differential operator $P(x, D) = \sum_j c_a(x) D^a$. Check how P interacts with a high frequency wave $u(x) = e^{2\pi i \langle \xi, x \rangle}$:

8.4 Suppose that $P(x, D) = \sum_a c_a(x) D^a$ is a linear differential operator, which we think of as a “polynomial in D ” with coefficients in x . Show that $P(x, D)u = P(x, \xi)u$, where $u(x) = e^{2\pi i \langle \xi, x \rangle}$ is a wave.



Suppose that $P(x, D)$ has degree k , and let $P_{\text{top}}(x, D) = \sum_{|a|=k} c_a(x) D^a$ be the highest order derivative terms. Ignoring lower order terms in momentum, $P(x, D)u = P_{\text{top}}(x, \xi)u + o(\xi)^k$. If we only want our wave to approximately solve the equation, to highest order for large momentum, near some point x , we need precisely $P_{\text{top}}(x, \xi) = 0$, a homogeneous polynomial equation in ξ for each point x . A *characteristic* of P at a point x is a solution ξ of $P_{\text{top}}(x, \xi) = 0$, representing the momentum of a wave which behaves roughly like a solution; the set of characteristics at a point x is the *characteristic variety*. Since $P_{\text{top}}(x, \xi)$ has all of its terms of the same degree k in ξ , if we rescale ξ , the k -th power of the scale factor just pops out of $P_{\text{top}}(x, \xi)$, so the characteristic variety is actually a cone passing through the origin.

8.5 If $P = \partial_t^2 - \partial_x^2 - \partial_y^2$ in x, y, t -space then write the momentum as a vector (ξ, η, τ) and check that the characteristic variety is $\tau^2 = \xi^2 + \eta^2$, a cone called the *light cone* representing momenta that produce waves consistent with the wave equation.

8.6 Write $\partial_t u$ as u_t and $\partial_x \partial_t u$ as u_{xt} , and so on. Find the characteristic variety of the *Euler–Tricomi equation* $u_{tt} = tu_{xx}$.

8.7 If $P = \partial_x^2 + \partial_y^2$, then the associated equation $Pu = 0$ is the equation of an electrostatic potential energy u , called the *Laplace equation*. Denote this differential operator as $\Delta = P(D)$, and call it the *Laplace operator*. Show that the characteristic variety is cut out by the equation $0 = \xi^2 + \eta^2$, which has $(\xi, \eta) = (0, 0)$ as its only solution.

Elliptic regularity

If the characteristic variety consists of $\xi = 0$ only, we say that the equation $Pu = 0$ is *elliptic*. Intuitively, an elliptic equation does not admit any high frequency wave solutions. Since a singular solution has high frequencies in its Fourier transform, we can expect that an elliptic equation doesn't admit any singular solutions.

For the moment, we restrict attention to linear operators of second order, so of the form

$$P = \sum_{ij} a_{ij}(x) D_i D_j + \sum_i a_i(x) D_i + a_0(x),$$

and with smooth coefficients $a_{ij}(x), a_i(x), a_0(x)$. Without loss of generality, we will assume that $a_{ij}(x) = a_{ji}(x)$.

8.8 Prove that P is elliptic just when, at each point x , the eigenvalues of the symmetric matrix $A(x) = (a_{ij}(x))$ are either all positive or all negative.

We assume from now on, without loss of generality, that all eigenvalues of $a_{ij}(x)$ are positive. The operator P is *uniformly elliptic* for $x \in U$ in some domain U if the eigenvalues of $A(x) = (a_{ij}(x))$ are bounded away from zero

throughout U . In other words, there is some constant bound $\lambda > 0$ so that $\sum_{ij} a_{ij}(x) \xi_i \xi_j \geq \lambda \|\xi\|^2$.

8.9 Prove that every linear second order elliptic differential operator is locally uniformly elliptic.

Theorem 8.1 (Elliptic regularity). *Suppose that P is a linear second order elliptic differential operator with smooth coefficients on a bounded domain U . If u is a distribution and $Pu \in L^2_k(U)$ for some integer $k \geq 0$, then $u \in L^2_{k+2}(U)$.*

This theorem is too difficult for us to prove; we will prove a weaker theorem.

Chapter 9

Pseudodifferential Operators

We define pseudodifferential operators and prove their basic properties.

Intuition

If $u \in C_c^\infty$ then $\mathcal{F}(D^a u) = \xi^a \hat{u}$, so $D^a u = \mathcal{F}^*(\xi^a \hat{u})$. Magic: the left hand side differentiates u , while the right hand side only involves integrals. Suppose that $p(x, \xi)$ is a polynomial in ξ , with coefficients smooth functions of x , say $p(x, \xi) = \sum_a c_a(x) \xi^a$. The smooth linear differential operator $p(x, D) = \sum_a c_a(x) D^a$ can be expressed in terms of integrals

$$\begin{aligned} p(x, D)u &= \sum_a c_a(x) \mathcal{F}^*(\xi^a \hat{u}), \\ &= \int p(x, \xi) \hat{u} e^{2\pi i \langle \xi, x \rangle} d\xi : \end{aligned}$$

any linear differential operator with smooth coefficients occurs as $p(x, D)$, so as such an integral. For example, the Laplace operator is

$$\Delta = \sum_j \partial_j \partial_j = -4\pi^2 \sum_j D_j D_j,$$

so

$$\Delta u = -4\pi^2 \int \|\xi\|^2 \hat{u} e^{2\pi i \langle \xi, x \rangle} d\xi.$$

Roughly, a pseudodifferential operator is anything given by the same sort of integral:

$$u \mapsto \int p(x, \xi) \hat{u} e^{2\pi i \langle \xi, x \rangle} d\xi,$$

but we might let $p(x, \xi)$ be a more general function than polynomial in ξ . We can write such an operator as $\mathcal{F}^* p(x, \xi) \mathcal{F}$. For example, the operator

$$u \mapsto \int \langle \xi \rangle^s \hat{u} e^{2\pi i \langle \xi, x \rangle} d\xi$$

maps $L_s^2(\mathbb{R}^n) \rightarrow L^2(\mathbb{R}^n)$ isometrically. Since each differential operator behaves like a polynomial in ξ , to invert a differential operator, we would expect to use an operator which behaves like the reciprocal of a polynomial in ξ .

If we try to solve $\Delta u = f$ for an unknown function u , we can try to take Fourier transform: $-4\pi^2 \|\xi\|^2 \hat{u} = \hat{f}$, so

$$\hat{u} = -\frac{1}{4\pi^2} \frac{\hat{f}}{\|\xi\|^2}.$$

This is fine away from $\xi = 0$, but near $\xi = 0$ small fluctuations in \hat{f} (representing small but very low frequency fluctuations in f) will cause huge fluctuations in \hat{u} , and so huge low frequency fluctuations in u . This makes the analysis tricky. Picture a smooth function $\Lambda(\xi)$ which is a little peak, equal to 1 near $\xi = 0$ and dies off to zero away from there. Let $V(\xi) = 1 - \Lambda(\xi)$, a little trough vanishing near $\xi = 0$ and equal to 1 farther away. (The letters Λ, V are chosen so that they look like a peak and a trough.) Then we could try to “approximate” the answer u by a function u_{approx} with

$$\hat{u}_{\text{approx}} = -\frac{V(\xi)}{4\pi^2} \frac{\hat{f}}{\|\xi\|^2}.$$

This won't give us an answer close to u in any sense, but in some respects it is an easier function to analyse, and the correction

$$\hat{u} - \hat{u}_{\text{approx}} = -\frac{\Lambda(\xi)}{4\pi^2} \frac{\hat{f}}{\|\xi\|^2}$$

is very low frequency, so $u - u_{\text{approx}}$ is not very small but is very smooth. We approximate so that we get the right singularities and fine small scale features, and then we need only make a very smooth and large scale correction.

Symbols

The function $p(x, \xi)$ is called the *symbol* of the pseudodifferential operator. In order to carry out analysis of pseudodifferential operators, we will need to restrict the possible symbols we allow. A *symbol of order s* on a domain $U \subset \mathbb{R}^n$ is a smooth function $p(x, \xi)$, defined for any $x \in U$ and $\xi \in \mathbb{R}^n$, so that for any compact set $K \subset U$ and any a, b , there are constants C_{ab} so that

$$|D_x^a D_\xi^b p(x, \xi)| \leq C_{ab} \langle \xi \rangle^{s-|b|},$$

for $x \in K$ and $\xi \in \mathbb{R}^n$. A sequence of symbols p_1, p_2, \dots of order s *converges* to zero, denoted $p_j \rightarrow 0$, if, for any a, b ,

$$\frac{|D_x^a D_\xi^b p_j(x, \xi)|}{\langle \xi \rangle^{s-|b|}} \rightarrow 0.$$

A sequence of symbols p_1, p_2, \dots of order s *converges* to a symbol p of order s , denoted $p_j \rightarrow p$, if $p_j - p \rightarrow 0$. If $p(x, \xi)$ is a symbol, write $p(x, D)$ to mean $\mathcal{F}^* p(x, \xi) \mathcal{F}$. The *order* of $p(x, D)$ means the order of its symbol $p(x, \xi)$.

9.1 Prove that the order of any differential operator with smooth coefficients is equal to its order as a pseudodifferential operator.

Applying pseudodifferential operators to poorly behaved functions

If $u \in \mathcal{S}$, we can differentiate under the integral sign in

$$p(x, D)u = \int p(x, \xi) e^{2\pi i \langle \xi, x \rangle} \left(\int u(y) e^{-2\pi i \langle \xi, y \rangle} dy \right) d\xi$$

by the dominated convergence theorem, any number of times, so $p(x, D)u \in C^\infty(U)$. The relation

$$x^a e^{2\pi i \langle \xi, x \rangle} = D_\xi^a e^{2\pi i \langle \xi, x \rangle}$$

implies

$$x^a p(x, D)u = (-1)^{|a|} \sum_{b+c=a} \mathcal{F}^* D_\xi^b p(x, \xi) \mathcal{F}(x^c u)$$

so that $p(x, D): \mathcal{S} \rightarrow \mathcal{S}$.

Lemma 9.1. *Suppose that $u \in \mathcal{S}$ and p is a pseudodifferential operator of finite order on the support of u . In case p might be complex valued, we write \bar{p} to mean the complex conjugate of p . The function*

$$\tilde{u}(\xi) = \int e^{-2\pi i \langle \xi, x \rangle} \bar{p}(x, \xi) u(x) dx$$

is Schwartz. The map $u \in C_c^\infty \mapsto \tilde{u} \in \mathcal{S}$ is continuous.

Proof.

$$\begin{aligned} \xi^a \tilde{u}(\xi) &= \xi^a \int e^{-2\pi i \langle \xi, x \rangle} \bar{p}(x, \xi) u(x) dx, \\ &= (-1)^{|a|} \int \left(D_x^a e^{-2\pi i \langle \xi, x \rangle} \right) \bar{p}(x, \xi) u(x) dx, \\ &= \int e^{-2\pi i \langle \xi, x \rangle} D_x^a (\bar{p}(x, \xi) u(x)) dx. \end{aligned}$$

If p has order s , then all of the $D_x^a \bar{p}(x, \xi)$ are dominated by $\langle \xi \rangle^s$, so

$$\begin{aligned} \langle \xi \rangle^k |\tilde{u}(\xi)| &= \langle \xi \rangle^k \left| \int e^{-2\pi i \langle \xi, x \rangle} \bar{p}(x, \xi) u(x) dx \right|, \\ &\leq C_k \langle \xi \rangle^s, \end{aligned}$$

i.e.

$$|\tilde{u}(\xi)| \leq C_k \langle \xi \rangle^{s-k}.$$

So \tilde{u} decays as rapidly as any rational function. Differentiating both sides of the definition

$$\tilde{u}(\xi) = \int e^{-2\pi i \langle \xi, x \rangle} \bar{p}(x, \xi) u(x) dx$$

we see that all derivatives of \tilde{u} have the same form as \tilde{u} does, and so they also decay faster than any rational function, so $\tilde{u} \in \mathcal{S}$.

Suppose that $u_j \rightarrow u \in \mathcal{S}$ and let $w_j = u - u_j$, so that $\langle \xi \rangle^k |\tilde{w}_j(\xi)|$ decay faster than any rational function. So then $\langle \xi \rangle^k \tilde{w}_j(\xi) \rightarrow 0$ by dominated convergence, uniformly on compact sets, and therefore uniformly because they decay, i.e. $\langle \xi \rangle^k \tilde{u}_j \rightarrow \langle \xi \rangle^k \tilde{u}$ uniformly, and similarly with any number of derivatives. \square

9.2 If $u, v \in C_c^\infty(U)$ on a domain U , prove that $\langle p(x, D)u, v \rangle = \langle \hat{u}, \tilde{v} \rangle$.

Define $p(x, D)u$ for $u \in \mathcal{S}'$, by $\langle p(x, D)u, v \rangle = \langle \hat{u}, \tilde{v} \rangle$, for any $v \in \mathcal{S}$, so $p(x, D): \mathcal{S}' \rightarrow \mathcal{S}'$. We henceforth discard the notation \tilde{v} .

The kernel

Formally, if we allowed ourselves to change the order of integration,

$$\begin{aligned} p(x, D)u &= \int \left(\int p(x, \xi) e^{2\pi i \langle \xi, x-y \rangle} d\xi \right) u(y) dy, \\ &= \int K(x, y) u(y) dy, \end{aligned}$$

where

$$K(x, y) = \int p(x, \xi) e^{2\pi i \langle \xi, x-y \rangle} d\xi.$$

If the order s of p is positive, this integral is meaningless. But if s is large enough negative (which of course never happens for a differential operator), then we can change the order of integration (by Fubini's theorem) as above, and K is a well-defined continuous function called the *kernel* of $p(x, D)$.

9.3 Suppose that $p(x, \xi) = p(\xi)$ is a symbol on \mathbb{R}^n which is independent of x and has order less than n . Explain why the kernel of $p(D)$ is $K(x, y) = \check{p}(x-y)$, the inverse Fourier transform. In particular, use this to find the kernel of the "Gaussian operator" $G(D)$ whose symbol is $G(\xi) = e^{-\pi \langle A\xi, \xi \rangle}$ where A is a symmetric matrix whose eigenvalues are positive.

More generally, no matter what the order, we define a distribution K on $U \times U$ by

$$\langle K, w \rangle = \iiint \bar{w}(x, y) p(x, \xi) e^{2\pi i \langle \xi, x-y \rangle} dy d\xi dx$$

for any $w \in C_c^\infty(U \times U)$ and call K the *kernel* of $p(x, D)$.

9.4 Suppose that $p_j(x, \xi) \rightarrow p(x, \xi)$ as symbols of some order s and that $p_j(x, D)$ has kernel $K_j(x, y)$ and $p(x, D)$ has kernel $K(x, y)$. (Note that K_j and K are only assumed to be distributions here.) Prove that $K_j \rightarrow K$, convergence in the sense of distributions.

9.5 Suppose that $P = p(x, D)$ is a pseudodifferential operator with kernel $K_P(x, y)$. Prove that the pseudodifferential operator Q with symbol $q(x, \xi) = \xi^a p(x, \xi)$ has kernel $K_Q(x, y) = D_y^a K_P(x, y)$.

9.6 Define a distribution $\delta(x - y)$ on test functions $w(x, y)$ by

$$\langle \delta(x - y), w \rangle = \int \bar{w}(x, x) dx.$$

Use the results of problem 9.3 and problem 9.4 in this chapter to prove that the kernel of the identity operator $Iu = u$ is $\delta(x - y)$. Use the results of problem 9.5 to show that the kernel of D^a is $K_{D^a}(x, y) = D_y^a \delta(x - y)$.

Lemma 9.2. *The kernel $K(x, y)$ of a pseudodifferential operator is a smooth function wherever $x \neq y$. If a pseudodifferential operator on a domain in \mathbb{R}^n has order s and kernel K then $(x - y)^a K(x, y)$ is C^k as long as $|a| > s + n + k$. So the order of a pseudodifferential operator is (up to adding a constant) the order of pole of its kernel.*

Proof. From the identity

$$(x - y)^a e^{2\pi i \langle \xi, x - y \rangle} = D_\xi^a e^{2\pi i \langle \xi, x - y \rangle}.$$

we see that

$$\begin{aligned} \langle (x - y)^a K, w \rangle &= \iiint p(x, \xi) \bar{w}(x, y) (x - y)^a e^{2\pi i \langle \xi, x - y \rangle} dy d\xi dx, \\ &= (-1)^{|a|} \iiint \bar{w}(x, y) e^{2\pi i \langle \xi, x - y \rangle} D_\xi^a p(x, \xi) dy d\xi dx, \end{aligned}$$

and if we make $|a|$ large enough then the distribution $(x - y)^a K$ is represented by the function

$$(-1)^{|a|} \int e^{2\pi i \langle \xi, x - y \rangle} D_\xi^a p(x, \xi) d\xi$$

which is differentiable as many times as we need if we make $|a|$ large enough. \square

A *smoothing operator* is an operator $P: \mathcal{S}' \rightarrow C^\infty$.

Lemma 9.3. *If a pseudodifferential operator $P = p(x, D)$ has order $-\infty$ (i.e. P has order s for all values of s), then P is smoothing.*

Proof. The kernel K of P is smooth, by lemma 9.2, so

$$Pu(x) = \int K(x, y) u(y) dy$$

is smooth by differentiation under the integral sign. \square

An operator P is *pseudolocal* if Pu is smooth on any open set on which u is smooth, for any distribution u .

Proposition 9.4. *Every pseudodifferential operator is pseudolocal.*

Proof. Suppose that u vanishes near some point x ; for simplicity take it to be the point $x = 0$, so $p(x, D)u(x) = \int K(x, y)u(y) dy$ is smooth near $x = 0$ because $K(x, y)$ is smooth for y away from x , while $u(y)$ vanishes for y near 0. More generally, if $u \in \mathcal{S}'$ is smooth near 0 then we can write $u = u_0 + u_1$ where u_0 is smooth with compact support and u_1 vanishes near 0, and then $p(x, D)u = p(x, D)u_0 + p(x, D)u_1$. \square

Asymptotic series

Roughly speaking, the big idea of pseudodifferential operators is to approximate operators, but not as a sum of a simple approximation and a small correction (as we would usually expect in analysis), but instead as a sum of a simple approximation and a smoothing operator.

If we have a pseudodifferential operator $p = p(x, D)$ and a sequence of pseudodifferential operators $p_1, p_2, \dots, p_j = p_j(x, D)$, we write $p \sim p_1 + p_2 + \dots$ to mean that the differences $p, p - p_1, p - p_1 - p_2, \dots$ have order going to $-\infty$; the formal sum $p_1 + p_2 + \dots$ is called an *asymptotic series* for p .

Proposition 9.5. *Suppose that p is a symbol of order s and let $p_\varepsilon(x, \xi) = p(x, \varepsilon\xi)$. Then $p_\varepsilon \rightarrow p_0$.*

Proof. Replace p by $p - p_0$ so that we can assume that $p_0 = 0$. It suffices to prove that

$$\langle \xi \rangle^{|b|-m} |D_x^a D_\xi^b p_\varepsilon(x, \xi)| \leq C_{ab} \varepsilon^m.$$

For $|b| = 0$, take a Taylor series in ε . For $|b| > 0$, this doesn't quite give the required power of ε , so we let the reader check that

$$\frac{\langle \xi \rangle^{|b|-m}}{\langle \varepsilon \xi \rangle^{|b|-m}} \leq \varepsilon^{m-|b|}$$

as $\varepsilon \rightarrow 0$. Therefore

$$\begin{aligned} \langle \xi \rangle^{|b|-m} |D_x^a D_\xi^b p_\varepsilon(x, \xi)| &\leq \langle \varepsilon \xi \rangle^{|b|-m} \varepsilon^{m-|b|} |\varepsilon^{|b|} D_x^a D_\xi^b p(x, \varepsilon\xi)| \\ &\leq \langle \varepsilon \xi \rangle^{|b|-m} \varepsilon^m |D_x^a D_\xi^b p(x, \varepsilon\xi)| \\ &\leq \langle \varepsilon \xi \rangle^{|b|-m} \varepsilon^m C_{ab} \langle \varepsilon \xi \rangle^{m-|b|}. \end{aligned}$$

\square

Theorem 9.6. *For any sequence p_1, p_2, \dots of symbols whose orders are finite and approach $-\infty$, there is a symbol p so that $p \sim p_1 + p_2 + \dots$. This symbol p is unique up to adding a smoothing operator, and doesn't change if we arbitrarily reorder the symbols p_1, p_2, \dots .*

Proof. Uniqueness: if we have two such, say p and q , their difference has order less than any of these p_j so is smoothing. The same idea works when we reorder.

Existence: Suppose that each of our symbols $p_j(x, \xi)$ is defined for x on some open set U and has order s_j , with $s_1, s_2, \dots \rightarrow -\infty$. Take a smooth function $\Lambda(\xi)$ so that $\Lambda = 1$ near the origin, while $\Lambda = 0$ everywhere far enough from the origin; let $V = 1 - \Lambda$. In the notation of proposition 9.5, $V_\varepsilon \rightarrow 0$ as a symbol of order 1, and therefore $V_\varepsilon p_j \rightarrow p_j$. Pick some numbers $\varepsilon_1, \varepsilon_2, \dots \rightarrow 0^+$ so that

$$|D_x^a D_\xi^b (V_{\varepsilon_j} p_j - p_j)| < \frac{\langle \xi \rangle^{s_j+1-|b|}}{2^j}.$$

Let $q_j = V_{\varepsilon_j} p_j$ and let $p = \sum q_j$. This sum is locally finite, because for j large enough we will get $\varepsilon_j \xi$ inside the locus where $V = 0$.

Since $p - \sum_{j < k} q_j$ is a sum of terms of the form $q_{k+1} + q_{k+2} + \dots$, we have to ensure that these terms decay like a convergent series, all of whose terms have large enough negative order, which they do:

$$|D_x^a D_\xi^b q_j| < \frac{\langle \xi \rangle^{s_j+1-|b|}}{2^j}.$$

Therefore the sum $p = \sum q_j$ is an asymptotic series. Each difference $q_j - p_j$ has order $-\infty$ so $p \sim p_1 + p_2 + \dots$ \square

Amplitudes

An adjoint P^* for an operator P means an operator so that $\langle P^* u, v \rangle = \langle u, P v \rangle$ for some dense collection of functions u, v . If we look for an adjoint for a pseudodifferential operator $P = p(x, D)$, we find

$$P^* v(x) = \int e^{2\pi i \langle \xi, x \rangle} \left(\int \bar{p}(y, \xi) v(y) e^{-2\pi i \langle \xi, y \rangle} dy \right) d\xi.$$

This is *not* a pseudodifferential operator in the sense above: if it were, then $\bar{p}(y, \xi)$ would have to depend only on x, ξ . We want a new definition which allows dependence on x, y, ξ : a *pseudodifferential operator* of order s on a domain $U \subset \mathbb{R}^n$ is a linear operator $A: C_c^\infty \rightarrow C^\infty$ of the form

$$Au(x) = \iint a(x, y, \xi) e^{2\pi i \langle \xi, x-y \rangle} u(y) dy d\xi$$

where the smooth function $a: U \times U \times \mathbb{R}^n$, called the *amplitude* has, for any compact set $K \subset U$, and any multiindices a, b, c , a constant $C_{a,b,c,K}$ so that on K

$$|D_x^a D_y^b D_\xi^c a(x, y, \xi)| \leq C_{a,b,c,K} \langle \xi \rangle^{s-|c|}.$$

If we let $b(x, y, \xi) = \overline{a(y, x, \xi)}$, one easily checks that the operator

$$A^* u(x) = \iint b(x, y, \xi) e^{2\pi i \langle \xi, x-y \rangle} u(y) dy d\xi$$

satisfies $\langle A^* u, v \rangle = \langle u, Av \rangle$ for all $u, v \in \mathcal{S}$.

Each pseudodifferential operator A with amplitude $a(x, y, \xi)$ has as *kernel* $K = K_A$ the distribution

$$\langle K, w \rangle = \iiint \bar{w}(x, y) a(x, y, \xi) e^{2\pi i \langle \xi, x-y \rangle} dy d\xi dx$$

for any $w \in C_c^\infty(U \times U)$.

9.7 Prove that the kernel $K(x, y)$ of a pseudodifferential operator is smooth away from $x = y$. Prove that if A is pseudodifferential operator of order s on a domain $U \subset \mathbb{R}^n$ then the kernel $K = K_A$ is C^k near $x = y$ as long as $0 > s + n + k$.

How do these complicated integrals Au involving amplitudes relate to the simpler integrals $p(x, D)u$ that we had before?

Theorem 9.7. *Any pseudodifferential operator A , say of order s and with amplitude $a(x, y, \xi)$, has the form $A = p(x, D) + \text{smoothing}$ where $p(x, \xi)$ is a symbol of order s with asymptotic series*

$$p(x, \xi) \sim \sum_a \frac{(2\pi i)^{|a|}}{a!} \partial_\xi^a \partial_y^a a(x, y, \xi)|_{y=x}.$$

Proof. Pick a smooth function $\Lambda(x, y)$ for $x, y \in U$ vanishing when x and y are close to one another and equal to 1 when x and y are far apart and let $V(x, y) = 1 - \Lambda(x, y)$. Let A_Λ be the operator with amplitude $\Lambda(x, y)a(x, y, \xi)$ and A_V be the operator with amplitude $V(x, y)a(x, y, \xi)$. If A has kernel $K(x, y)$, then A_Λ has kernel $\Lambda(x, y)K(x, y)$, while A_V has kernel $V(x, y)K(x, y)$. Since $V(x, y)$ vanishes near $x = y$, A_V is a smoothing operator. Clearly $K_V(x, y)$ vanishes near $x = y$ so has no singularity. Problem 9.7 tells us that A_V has order $-\infty$, while $K_\Lambda(x, y) = K(x, y)$ near $x = y$, so A_Λ has the same order as A . It thus suffices to prove the theorem for A_Λ rather than for A : we can assume from now on that $a(x, y, \xi) = 0$ if x and y are far enough apart.

Although Au is only defined for $u \in \mathcal{S}$, we can define Au for $u \in C^\infty$ to mean $Au = \int_U K(x, y)u(y)$, because once we force x to lie in some compact set, $K(x, y) = \Lambda(x, y)K(x, y)$ vanishes for y outside some compact set, so $K(x, y)u(y)$ is smooth with compact support in y . Define

$$p(x, \xi) = e^{-2\pi i \langle \xi, x \rangle} A e^{2\pi i \langle \xi, x \rangle}.$$

Pick any $u \in \mathcal{S}$ and then $\hat{u} \in \mathcal{S}$ and write

$$u(x) = \int \hat{u} e^{2\pi i \langle \xi, x \rangle}.$$

This integral is a limit of Riemann sums, by the smoothness and rapid decay of the integrand. Clearly $A: \mathcal{S} \rightarrow \mathcal{S}$ is continuous, as it is just

$$Au(x) = \int a(x, y, \xi) e^{2\pi i \langle \xi, x-y \rangle} u(y) dy d\xi.$$

So we can write u as a limit of Riemann sums above, and use this to interchange the integration in

$$\begin{aligned} Au(x) &= \int a(x, y, \xi) e^{2\pi i(\langle \xi, x-y \rangle + \langle \eta, y \rangle)} \hat{u}(\eta) d\eta dy d\xi, \\ &= \int a(x, y, \xi) e^{2\pi i(\langle \xi, x-y \rangle + \langle \eta, y \rangle)} \hat{u}(\eta) dy d\xi d\eta, \\ &= \int A(e^{2\pi i\langle \eta, y \rangle}) \hat{u}(\eta) d\eta \\ &= p(x, D)u. \end{aligned}$$

Expanding out:

$$p(x, \xi) = \int a(x, y, \eta) e^{2\pi i\langle \eta - \xi, x - y \rangle} dy d\eta.$$

It is convenient to change variables to $z = y - x$:

$$p(x, \xi) = \int a(x, x + z, \eta) e^{-2\pi i\langle \eta - \xi, z \rangle} dz d\eta.$$

Let

$$b(x, z, \xi) = a(x, x + z, \xi),$$

and let $\hat{b}(x, \eta, \xi)$ be the Fourier transform of $b(x, z, \xi)$ in the z -variable:

$$\begin{aligned} \hat{b}(x, \eta, \xi) &= \int b(x, z, \xi) e^{-2\pi i\langle \eta, z \rangle} dz, \\ &= \int a(x, x + z, \xi) e^{-2\pi i\langle \eta, z \rangle} dz. \end{aligned}$$

Therefore

$$p(x, \xi) = \int \hat{b}(x, \eta, \eta + \xi) d\eta.$$

If we force x to stay in some compact set, $b(x, y, \xi)$ vanishes for y outside some larger compact set.

Since we know that a has order s , we see that

$$\left| D_x^a D_\xi^c \hat{b}(x, \eta, \xi) \right| \leq C_{ac} \langle \xi \rangle^{s-|c|}.$$

Since $a(x, x + z, \xi)$ is C_c^∞ in z for fixed x , we know that $\hat{b}(x, \eta, \xi)$ is Schwartz in η , and the various Schwartz estimates are uniform in x and ξ . We expand $a(x, x + z, \xi)$ in a Taylor series in z , plug in to express b as a Taylor series, and expand out p , and check that we have the required estimates. \square

For example, if $p(x, \xi)$ is a symbol of order s , then its adjoint $p^*(x, \xi)$ is a symbol of the same order with asymptotic series

$$p^*(x, \xi) \sim \sum_a \frac{(2\pi i)^a}{a!} \partial_\xi^a \partial_y^a \overline{p(y, \xi)}.$$

9.8 On \mathbb{R}^2 let $p(x, D) = x_1 D_1$ and $q(x, D) = x_1 D_2$. Compute $r(x, D) = p(x, D)q(x, D)$. Show that $r(x, \xi) \neq p(x, \xi)q(x, \xi)$.

Theorem 9.8. *If $p(x, D)$ is a pseudodifferential operator of order s_p and $q(x, D)$ is a pseudodifferential operator of order s_q then $r(x, D) = p(x, D)q(x, D)$ is a pseudodifferential operator of order $s_r = s_p + s_q$ with asymptotic series*

$$r(x, \xi) = \sum_a \frac{(2\pi i)^{|a|}}{a!} D_\xi^a p(x, \xi) D_x^a q(x, \xi).$$

Proof. Let

$$r(x, \xi) = \iint p(x, \eta) q(y, \xi) e^{2\pi i \langle \eta - \xi, x - y \rangle} dy d\eta.$$

Check that, if we can justify an interchange of integrals, we have $p(x, D)q(x, D) = r(x, D)$. To justify the interchange of integrals, we need to use the same idea as in theorem 9.7 to replace $p(x, \xi)$ and $q(x, \xi)$ with amplitudes $a(x, y, \xi)$ and $b(x, y, \xi)$ which vanish when x is not close to y . \square

Elliptic regularity

A *parametrix* for a pseudodifferential operator $p(x, D)$ is a pseudodifferential operator $q(x, D)$ so that both $p(x, D)q(x, D)$ and $q(x, D)p(x, D)$ differ from I by pseudodifferential operators of order $-\infty$; essentially $q(x, D)$ is an inverse of $p(x, D)$.

A pseudodifferential operator $p(x, D)$ of order s on an open set $U \subset \mathbb{R}^n$ is *elliptic* if, for any compact set $K \subset U$, there is a constant $C_K > 0$ so that $|p(x, \xi)| \geq C_K |\xi|^s$.

Theorem 9.9. *Every elliptic pseudodifferential operator has a parametrix.*

Proof. Pick any smooth function $\Lambda(\xi)$ equal to 1 near 0 and equal to 0 far enough from 0 and let $V(\xi) = 1 - \Lambda(\xi)$. Let $q_0(x, \xi) = V(\xi)p(x, \xi)^{-1}$; by the chain rule q_0 is a symbol of order $-s$. By theorem 9.8,

$$\begin{aligned} p(x, D)q_0(x, D) &= I + r_1(x, D), \\ q_0(x, D)p(x, D) &= I + r_2(x, D), \end{aligned}$$

for some $r_1(x, D), r_2(x, D)$ symbols of order -1 . By theorem 9.6, there is a pseudodifferential operator $s(x, D)$ of order -1 so that

$$I + s(x, D) \sim I - r_2(x, D) + r_2(x, D)^2 - r_2(x, D)^3 + \dots$$

Let $q(x, D) = (I + s(x, D))q_0(x, D)$ and then

$$q(x, D)p(x, D) = I + t(x, D)$$

where $t(x, D)$ has order $-\infty$. Similarly construct a pseudodifferential operator $q'(x, D)$ so that

$$p(x, D)q'(x, D) = I + t'(x, D)$$

where $t'(x, D)$ has order $-\infty$. Expand

$$(q(x, D)p(x, D))q'(x, D) = q(x, D)(p(x, D)q'(x, D))$$

to show that $q = q'$ up to an error of order $-\infty$. \square

Theorem 9.10 (Elliptic regularity). *For any elliptic pseudodifferential operator $p(x, D)$, if $f \in C^\infty$ and $Pu = f$ for some $u \in \mathcal{S}'$ then $u \in C^\infty$.*

Proof. Take a parametrix $q(x, D)$, so that $q(x, D)p(x, D) = I + r(x, D)$ where $r(x, D)$ is a smoothing operator. Apply $q(x, D)$ to both sides of $p(x, D)u = f$ to get $u = q(x, D)f - r(x, D)u$, which is clearly smooth. Therefore if u exists, then it is smooth. \square

Theorem 9.11 (Elliptic local solvability). *For any elliptic pseudodifferential operator $p(x, D)$, if $f \in C^\infty$, then the equation $Pu = f$ admits local solutions $u \in C^\infty$ for any $f \in C^\infty$.*

Proof. Take a parameterix $q(x, D)$, so that $q(x, D)p(x, D) = I + r(x, D)$ where $r(x, D)$ is a smoothing operator. The smoothing operator $r(x, D)$ has a smooth kernel, say $R(x, y)$, so

$$r(x, D)u(x) = \int R(x, y)u(y) dy.$$

We suppose that our operators are defined on functions on some open set $U \subset \mathbb{R}^n$, and then we choose a relatively compact domain $V \subset U$. We then consider the map $r(x, D): C^\infty(V) \rightarrow C^\infty(V)$; the Hölder inequality gives

$$\begin{aligned} \|r(x, D)u\|_{L^2(V)}^2 &= \int_V |r(x, D)u(x)|^2 dx, \\ &= \int_V \left| \int_V R(x, y)u(y) dy \right|^2 dx, \\ &\leq M \|u\|_{L^2(V)}^2 \end{aligned}$$

where

$$\begin{aligned} M &= \int_V \max_{y \in V} |R(x, y)|^2 dx, \\ &\leq \text{Vol}(V) \max_{x, y \in V} |R(x, y)|^2. \end{aligned}$$

If we make V small enough, then M becomes as small as we like. In particular, we can arrange that $M < 1$, and then the operator $r(x, D)$ is bounded on $L^2(V)$ with norm at most M . Therefore the sum $(I + r(x, D))^{-1} = I - r(x, D) +$

$r(x, D)^2 + \dots$ converges to a bounded operator on $L^2(V)$. This operator is given by a convergent sum of integrals with smooth kernels; one easily sees that (if we make V small enough) this operator preserves smoothness as well by writing out the terms and differentiating.

Take any smooth function f on V , extend to all of U smoothly, and define $u = (I + r(x, D))^{-1}q(x, D)f$. This smooth function then satisfies $(I + r(x, D))u = q(x, D)f$, i.e. $u = q(x, D)f - r(x, D)u$. We then apply $p(x, D)$ to both sides to find $p(x, D)u = f$. \square

Bibliography

- [1] Terence Tao, *An introduction to measure theory*, Graduate Studies in Mathematics, vol. 126, American Mathematical Society, Providence, RI, 2011. MR 2827917 (2012h:28003) 4
- [2] Richard L. Wheeden and Antoni Zygmund, *Measure and integral*, Marcel Dekker Inc., New York, 1977. 4

List of notation

\mathbb{C}^n	the set of all n -tuples of complex numbers, 1
\mathbb{R}^n	the set of all n -tuples of real numbers, 1
$f(S)$	image of set S via the map f , 1
$f^{-1}T$	preimage of set T via the map f , 1
$[a, b]$	closed interval from a to b , 2
$B_r(x)$	open ball of radius r about the point x , 2
$\bar{B}_r(x)$	closed ball of radius r about the point x , 2
C^∞	smooth functions, 3
C^k	k times continuously differentiable functions, 3
\bar{S}	closure of a set S , 2
∂^a	multiindex partial derivative, 3
∂_i	$\frac{\partial}{\partial x_i}$, 2
∂_x	$\frac{\partial}{\partial x}$, 2
∂S	boundary of a set S , 2
$a!$	multiindex factorial, 3
df	differential, 2
$o(f(x))$	something small relative to $f(x)$, 3
x^a	multiindex power, 3
$C^{k,\alpha}$	Hölder continuity of order k, α , 3
$\ f\ _{C^k}$	C^k norm: sup of derivatives of order $\leq k$, 3
C_c^∞	test functions, 4
$L^\infty(X)$	the set of bounded measurable functions on X , 6
$L^p(X)$	the set of p -power integrable functions on X , 6
$\langle f, g \rangle$	L^2 inner product, 6
$\ f\ _\infty$	uniform norm, 6
$\ f\ _{L^p}$	L^p -norm, 6
$L_{\text{loc}}^p(U)$	locally L^p functions on an open set U , 7
$f * g$	convolution, 13
$L_k^p(U)$	Sobolev space, 23

λX	sensitivity to large humps of a function space X , 25
σX	sensitivity to small bumps of a function space X , 25
ϕX	sensitivity to high frequencies of a function space X , 26
\mathcal{S}	Schwartz functions, 29
\hat{f}	Fourier transform, 30
\mathcal{F}	Fourier transform, 32
\check{f}	inverse Fourier transform, 34
\mathcal{F}^*	inverse Fourier transform, 34
δ	Dirac delta function, 37
$\langle x \rangle$	Japanese bracket, 41
L_k^2	Sobolev L^2 space, 41
$\text{tr}_X(f)$	trace (i.e. restriction) of a function f to a set X , 43
D	$\partial/2\pi i$, 57
Δ	Laplace operator, 58

Index

- action, **48**
- almost everywhere, **5**
- amplitude, **67**
- asymptotic series, **66**
- ball
 - closed, **2**
 - open, **2**
- Banach space, **9**
- bell curve, **16**
- best constant, **26**
- boundary
 - of a set, **2**
- bounded, **2**
- box, **2**
- Cauchy sequence, **9**
- characteristic, **58**
- characteristic variety, **58**
- compact, **2**
- complex Euclidean space, **1**
- convergence
 - of distributions, **39**
 - of tempered distributions, **40**
- derivative
 - strong, **20**
 - weak, **20**
- Dirac delta function, **37**
- distance, **2**
- distribution, **37**
 - tempered, **40**
- domain, **2**
- dominated convergence theorem, **5**
- elliptic, **58**
 - pseudodifferential operator, **70**
 - uniformly, **58**
- embedded subspace, **26**
 - compactly, **26**
- Euclidean space, **1**
- Euler–Tricomi equation, **58**
- flow line, **56**
- Fourier transform, **30**
- function
 - Schwartz, **29**
- Gaussian, **16**
- Heaviside function, **22**
- high frequency sensitivity, **26**
- hypersurface measure, **8**
- image, **1**
- indicator function, **11**
- integrable
 - Lebesgue, **5**
 - Riemann, **5**
- Japanese bracket, **41**
- kernel
 - pseudodifferential operator, **64**
- Laplace equation, **58**
- Laplace operator, **58**
- large hump
 - sensitivity, **25**
- length, **2**
- light cone, **58**
- locally integrable, **7**
- measure, **5**
- measureable
 - function, **5**
 - set, **5**
- metric space
 - complete, **9**
- momentum, **57**
- New York function, **11**

- open cover, [4](#)
- open set, [2](#)
- order
 - pseudodifferential operator, [62](#)
- outer measure, [5](#)
- parametrix, [70](#)
- partition of unity, [4](#)
- positive definite, [55](#)
- preimage, [1](#)
- pseudodifferential operator, [67](#)
 - elliptic, [70](#)
- pseudolocal, [65](#)
- rapidly decreasing, [29](#)
- Riemann integral, [5](#)
- Schwartz function, [29](#)
- sensitivity, [25](#)
 - to high frequencies, [26](#)
 - to large humps, [25](#)
 - to small bumps, [25](#)
- small bump
 - sensitivity, [25](#)
- smooth, [3](#)
- smoothing operator, [65](#)
- Sobolev norm, [23](#)
- Sobolev space, [23](#)
- strong derivative, [20](#)
- subordinate, [4](#)
- support, [4](#), [38](#)
- surface, [7](#)
- symbol, [62](#)
- tangent hyperplane, [8](#)
- tangent vector, [8](#)
- Taylor series, [3](#)
- tempered
 - distribution, [40](#)
- test function, [4](#)
- theorem
 - dominated convergence, [5](#)
 - Kondrashov–Rellich compactness, [27](#)
 - Sobolev embedding, [27](#)
- vector field, [55](#)
- weak convergence, [9](#)
- weak derivative, [20](#)
- weak limit, [9](#)
- width, [47](#)