# Linear Analysis and Partial Differential Equations

# Lecture Notes

University of Illinois
at Urbana–Champaign

Richard S. Laugesen *

December 22, 2020

## Preface

A *textbook* presents far more material than any professor can cover in class. These *lecture notes* present exactly what I covered during the one semester course Linear Analysis and Partial Differential Equations (Math 554) at the University of Illinois, Urbana–Champaign, in Fall 2020. A few enhancements were added after the semester notably the 1-dimensional case of the Sobolev inequalities and Rellich–Kondrachev compactness, in Chapter 3.

The exercises interspersed throughout the notes were covered on homework assignments, and were generally not stated in class.

I would welcome corrections, and suggested improvements.

Richard S. Laugesen EMAIL: Laugesen@illinois.edu
Department of Mathematics, University of Illinois Urbana–Champaign, U.S.A.

## Introduction

This course develops the modern theory of partial differential equations on a bounded domain, concentrating on second order equations of elliptic, parabolic and hyperbolic type. These equations can rarely be solved explicitly, and so the task is to develop abstract solution methods that enable us to prove well-posedness: existence, uniqueness and continuous dependence of solutions on the data. Along the way we will glimpse how these solutions might be approximated numerically, because our solution methods typically proceed by constructing a convergent sequence of approximate solutions.

After proving well-posedness, we examine qualitative properties of solutions, most notably maximum principles for elliptic and parabolic PDEs, and finite speed of propagation for hyperbolic PDEs.

While our focus in this course is primarily linear, we build a foundation for studying the nonlinear world. The course begins with fixed point methods for contractions and compact operators, as a way of introducing fundamental ideas of functional analysis in connection with nonlinear ODEs. The course concludes with semigroup methods, which allow us at the very end to apply fixed point methods as a tool for solving nonlinear PDEs.

*Style of the course.* Our scheme in each chapter is to introduce ideas from functional analysis and apply them to obtain results on partial differential equations. The applications require increasing amounts of hard analysis, as the course proceeds. PDE theory is not a subset of functional analysis!

# Contents

# Prerequisites and notation

We assume familiarity with metric space topology: Cauchy sequences, completeness, continuity, and compactness both in terms of open covers and in terms of subsequences (sequential compactness).

The basics of functional analysis are also assumed to be known.

All functions are assumed to be measurable. The essential function spaces are:

$$\begin{aligned} C(X) &= \text{continuous functions on the topological space } X, \\ L^p(X, \mu) &= p\text{-th power integrable functions on the measure space } (X, \mu), \\ L^\infty(X, \mu) &= \text{essentially bounded functions on } (X, \mu). \end{aligned}$$

The latter two spaces become Banach spaces under the standard $L^p$ and $L^\infty$ norms, and $C(X)$ is a Banach space under the max-norm $\|T\| = \max_{x \in X} |T(x)|$ provided the topological space $X$ is compact.

The $\ell^p$-space is simply $L^p$ with a counting measure.

Functions will always be real-valued, unless we say otherwise.

Superscripts can indicate iteration of a mapping ($T^3 = T \circ T \circ T$) or else powers of a function or number ($f^3 = fff$). The meaning will be clear from the context.

The Lebesgue measure of a set $E \subset \mathbb{R}^N$ is denoted $|E|$.

ONB means orthonormal basis.

On a domain $U \subset \mathbb{R}^N$, for $N \geq 1$:

$$\begin{aligned} C(U) &= \{\text{continuous functions on } U\} \\ C(\overline{U}) &= \{\text{uniformly continuous functions on } U\} \\ C^k(U) &= \{k\text{-times differentiable functions on } U \\ &\qquad\qquad \text{whose derivatives are continuous}\} \\ C^k(\overline{U}) &= \{k\text{-times differentiable functions on } U \\ &\qquad\qquad \text{whose derivatives are uniformly continuous}\} \end{aligned}$$

If $u \in C^k(\overline{U})$ then each derivative of $u$ up to order $k$ extends to a continuous function on $\overline{U}$, by an exercise in real analysis.

The notation $U \Subset V$ means $U$ is compactly contained in $V$, in other words, that $\overline{U}$ is a compact subset of $V$.

# Day-by-day plan (for 50 minute classes)

*Day 1*

1.1 Contraction mapping principle
1.2 Application to Picard's existence and uniqueness theorem

*Day 2*

1.2 Application to Picard's existence and uniqueness theorem, cont.
1.3 Fredholm equation (statement only)
1.4 Contraction-power mapping principle

*Day 3*

1.5 Application to Volterra's equation
1.6 Compact operators (definition, and example of nonlinear integral operator)

*Before the next class*

Read the first three pages of Chapter 2 (up through Example 2.8), to refresh your memory on the basics of Hilbert space theory. We will not cover this material in class.

*Day 4*

1.6 Compact operators (approximation by finite-rank operators)
1.7 Brouwer and Schauder fixed point theorems

*Day 5*

1.8 Application to Peano's existence theorem
2.1 Hilbert space basics (noncompleteness of $C^1$ with Sobolev inner product)

*Day 6*

2.1 Hilbert space basics (Orthogonal Decomposition, no proof)
2.1 Hilbert space basics (Riesz Representation)

*Day 7*

2.2 Weak solution of Poisson equation
2.3 Orthonormal bases

*Day 8*

2.3 Orthonormal bases, cont.
2.4 Weak compactness

*Day 9*

2.4 Weak compactness, cont.
3.1 Green's theorem, integration by parts [reading; not covered in class]
3.2 Mollification and smoothing

*Day 10*

3.2 Mollification and smoothing, cont.

*Day 11*
3.3 Weak derivatives and Sobolev spaces
*Day 12*
3.4 Approximating Sobolev functions by smooth functions
3.5 Sobolev functions vanishing on the boundary
*Day 13*
3.6 Extending past the boundary
3.7 Boundary traces (statement)
*Day 14*
3.7 Boundary traces (proof)
3.8 Sobolev inequalities
*Day 15*
3.8 Sobolev inequalities, cont.
*Day 16*
3.8 Sobolev inequalities, cont.
*Day 17*
3.9 Compact imbedding of Sobolev spaces (Rellich-Kondrachev)
*Day 18*
3.9 Compact imbedding of Sobolev spaces, cont.
3.10 Application: Poisson's Equation via Calculus of Variations
*Day 19*
3.10 Application: Poisson's Equation via Calculus of Variations, cont.
4.2 Spectral theorem for compact, selfadjoint operators (statement only)
*Day 20*
4.1 Abstract spectral theory for sesquilinear forms
*Day 21*
4.1 Abstract spectral theory for sesquilinear forms, cont.
*Day 22*
4.3 Application to elliptic operators
*Day 23*
5.1 Generalized Poisson equation - wellposedness
*Day 24*
5.1 Generalized Poisson equation, cont.
5.2 Regularity of solutions
*Day 25*
5.2 Regularity of solutions, cont.
*Day 26*
5.2 Regularity of solutions, cont.

5.3 Weak Maximum Principles

*Day 27*

5.3 Weak Maximum Principles, cont.

5.4 Strong Maximum Principles

*Day 28*

5.4 Strong Maximum Principles, cont.

4.4 Lax–Milgram and nonsymmetric sesquilinear forms (and connections to non-selfadjoint elliptic PDEs)

*Day 29*

Leeway

*Day 30*

6.1, 6.2 Parabolic equations and the Galerkin approximation

*Day 31*

6.2 cont. and 6.3 Parabolic equations and the Galerkin approximation

*Day 32*

6.4 Energy estimates and weak solutions

*Day 33*

6.5 Maximum principles

*Day 34*

7.1, 7.2 Hyperbolic equations

*Day 35*

7.3 Finite speed of propagation

*Day 36*

8.1 Generators, Resolvents

*Day 37*

8.1 Generators, Resolvents, cont.

*Day 38*

8.2 Statement of Hille-Yosida, and sketch of proof

*Day 39*

8.3 Dissipative operators

*Day 40*

8.3 Dissipative operators, cont.

*Day 41*

8.4 Application: solving parabolic and hyperbolic PDEs by semigroups

*Day 42*

8.5 Application: nonhomogeneous and nonlinear evolution equations

# Chapter 1

# Fixed point theorems and applications to ODEs

**References** [Chicone, Zeidler]

**Notation** $(X, d)$ is a metric space, throughout this chapter, and $T : X \to X$ is a mapping. The letters $m$ and $n$ denote nonnegative integers, unless specified otherwise.

We begin with contraction mapping principles. As an application we prove Picard's theorem on existence of solutions for nonlinear ODEs, by inverting the differential equation into an integral equation and then solving the integral equation.

The same pattern recurs later in the course for PDEs, where one proves existence of a solution by inverting in some fashion to obtain an integral equation. At heart, this approach relies on the principle that differentiation is a "bad" operator while integration (the inverse of differentiation) is a "good" operator.

If you have never considered this principle, then examine the following simple example in $C[0, 1]$ with the max-norm. The function $x^n$ has norm $1$ while its derivative $nx^{n-1}$ has norm $n$ and its integral $x^{n+1}/(n+1)$ has norm $1/(n+1)$. Let $n \to \infty$ and observe that the norm of the derivative blows up while that of the integral is well behaved (and even decays to zero).

## 1.1 Contraction mapping principle

**Definition 1.1.** A **fixed point** of $T$ is an element $x_* \in X$ such that $Tx_* = x_*$. The fixed point is called **globally attracting** if every sequence of iterates converges to the fixed point ($T^n x \to x_*$ as $n \to \infty$ for each $x \in X$).

*Example* 1.2. The map $T(x) = x/2$ has a globally attracting fixed point $x_* = 0$. The map $T(x) = x^3$ from $\mathbb{R}$ to $\mathbb{R}$ has fixed points $0, 1$ and $-1$, none of which is globally attracting (why?).

**Definition 1.3.** $T$ is a **contraction** if for some $\alpha \in [0, 1)$ one has

$$d(Tx, Ty) \leq \alpha d(x, y), \qquad x, y \in X.$$

That is, $T$ is Lipschitz continuous with Lipschitz constant less than $1$. In particular, contractions are continuous mappings.

**Theorem 1.4** (Banach fixed point theorem, or contraction mapping principle)**.** *Assume the metric space is complete and* $T$ *is a contraction. Then* $T$ *has a unique fixed point* $x_*$. *This fixed point is globally attracting with geometric rate of convergence:*

$$d(T^n x, x_*) \leq \frac{\alpha^n}{1-\alpha} d(Tx, x), \qquad x \in X, \quad n \geq 0.$$

Incidentally, if the contraction depends smoothly on a parameter, then the fixed point depends smoothly on the parameter too [Chicone, Section 1.9.3].

*Proof.* Step 1— Uniqueness. Suppose $x_*$ and $y_*$ are fixed points, so that $Tx_* = x_*$ and $Ty_* = y_*$. The definition of a contraction then implies that $d(x_*, y_*) = d(Tx_*, Ty_*) \leq \alpha d(x_*, y_*)$, and so $(1-\alpha)d(x_*, y_*) \leq 0$. Dividing by $1-\alpha$ implies $d(x_*, y_*) \leq 0$ and hence $x_* = y_*$.

Step 2 — Each sequence of iterates is Cauchy. Now consider an arbitrary point $x \in X$ and consider the sequence of iterates $T^n x, n \geq 0$. The contraction property implies that

$$d(T^{n+1}x, T^n x) \leq \alpha d(T^n x, T^{n-1}x) \leq \cdots \leq \alpha^n d(Tx, x), \tag{1.1}$$

which means the distance between consecutive iterates decays geometrically.

To estimate distances between nonconsecutive iterates we invoke the triangle inequality with intermediate iterates and call on the preceding estimate: for $m \geq 1$,

$$\begin{aligned} d(T^{n+m}x, T^n x) &\leq d(T^{n+m}x, T^{n+m-1}x) + \cdots + d(T^{n+1}x, T^n x) \\ &\leq (\alpha^{n+m-1} + \cdots + \alpha^n) d(Tx, x) \qquad \text{by (1.1)} \\ &= \alpha^n \frac{1-\alpha^m}{1-\alpha} d(Tx, x) \\ &\leq \frac{\alpha^n}{1-\alpha} d(Tx, x). \end{aligned} \tag{1.2}$$

The right side can be made arbitrarily small by taking $n$ large, which implies that the sequence $\{T^n x\}$ is Cauchy. The metric space is complete, and so the sequence has a limit $x_* = \lim_{n\to\infty} T^n x$.

Step 3 — $x_*$ is a fixed point, because

$$Tx_* = T \lim_{n\to\infty} T^n x = \lim_{n\to\infty} T^{n+1} x = x_* \tag{1.3}$$

where continuity of the contraction justifies our taking $T$ inside the limit.

Uniqueness of fixed points now implies that $T^n x \to x_*$ for all $x$, so that the fixed point is globally attracting.

Step 4 — Convergence estimate. Letting $m \to \infty$ in (1.2) yields the convergence estimate in the theorem. □

***Exercise*** 1.1. Fix $a > 0$. Prove that a unique continuous function $f \in C[-a, a]$ exists solving

$$f(x) = 1 + \frac{1}{\pi} \int_{-a}^{a} \frac{1}{1 + (x-y)^2} f(y)\, dy, \qquad x \in [-a, a].$$

Show also that the solution $f$ is nonnegative.
*Note.* This integral equation arises in a quantum mechanical model of gas particle motion in one dimension (the Lieb–Liniger model).

If $T$ is known only to be contracting on a subset $Y \subset X$ (meaning $T$ maps $Y$ into itself and $T$ is a contraction on $Y$), then the fixed point theorem still applies provided $Y$ is closed, because then $(Y, d)$ is a complete metric space in its own right.

The next lemma gives a sufficient condition that insures a contracting $T$ will map a closed ball into itself (and thus have a fixed point there, provided $X$ is complete).

**Lemma 1.5.** *Let* $x_0 \in X$ *and* $r > 0$*. Suppose* $T : X \to X$ *has the contraction property on the closed ball* $Y = \{x : d(x, x_0) \le r\}$*, meaning there exists* $\alpha \in [0, 1)$ *such that*

$$d(Tx, Ty) \le \alpha d(x, y), \qquad x, y \in Y.$$

*If* $d(Tx_0, x_0) \le (1 - \alpha)r$*, then* $T(Y) \subset Y$*.*

***Exercise*** 1.2. Prove Lemma 1.5.

## 1.2 *Application:* Picard's theorem for first order ODEs

Consider the first order ODE initial value problem

$$x'(t) = f\big(t, x(t)\big), \qquad x(t_0) = x_0, \tag{1.4}$$

where $t_0, x_0 \in \mathbb{R}$ are given. Assume $f$ is a continuous function on the closed rectangle

$$R = \{(t, x) : |t - t_0| \le a, |x - x_0| \le b\}$$

for some $a, b > 0$, and that $f$ is Lipschitz in the $x$-variable:

$$|f(t, x) - f(t, y)| \le K|x - y| \qquad \text{for all } (t, x), (t, y) \in R,$$

for some $K > 0$. Write $c = \max |f|$, and let $\tau < \min\{a, b/c, 1/K\}$.

**Theorem 1.6** (Picard)**.** *The initial value problem* (1.4) *has a unique solution* $x(\cdot)$ *on the interval* $[t_0 - \tau, t_0 + \tau]$*.*

*Proof.* Take $t_0 = 0$ for notational simplicity, so that the interval is $J = [-\tau, \tau]$.

Step 1 — Reformulating as an integral operator fixed point problem. We will prove existence of a unique solution to the *integrated* form of the initial value problem, which is

$$x(t) = x_0 + \int_0^t f\big(s, x(s)\big)\, ds, \qquad t \in J. \tag{1.5}$$

***Exercise*** 1.3. In this exercise $f$ is continuous (but not necessarily Lipschitz continuous in the second variable).
(i) Prove that if a continuous function $x(\cdot)$ satisfies the integrated form (1.5), then it is differentiable and satisfies (1.4) on $J$.
(ii) In the reverse direction, show that any differentiable function $x(\cdot)$ satisfying (1.4) on $J$ must have continuous derivative and hence can be integrated to obtain (1.5).

Consider the Banach space $C(J)$ of continuous functions on $J$ with the max-norm $\|x\| = \max_{t\in J} |x(t)|$. We will work on the closed ball of radius $c\tau$ centered at the constant function $x_0$; call this ball

$$Y = \{x \in C(J) : \max_{t\in J} |x(t) - x_0| \le c\tau\}.$$

If $x \in Y$ then $(s, x(s)) \in R$ for all $s \in J$, since $c\tau < b$ by definition of $\tau$. Hence it makes sense to define a map $T : Y \to C(J)$ by

$$(Tx)(t) = x_0 + \int_0^t f\big(s, x(s)\big)\, ds, \qquad t \in J.$$

The integrated form (1.5) then says that

$$x = Tx$$

as functions in $C(J)$. In other words, to solve the ODE we must find a fixed point of the mapping $T$.

Step 2 — $T$ maps the closed ball into itself: if $x \in Y$ then $Tx \in Y$ because

$$\begin{aligned} |Tx(t) - x_0| &= \Big| \int_0^t f\big(s, x(s)\big)\, ds \Big| \\ &\le |t| c \le c\tau \end{aligned}$$

for all $t \in J$.

Step 3 — $T$ contracts the closed ball: if $x, y \in Y$ then for all $t \in J$,

$$\begin{aligned} |Tx(t) - Ty(t)| &= \Big| \int_0^t [f\big(s, x(s)\big) - f\big(s, y(s)\big)]\, ds \Big| \\ &\le \tau K \max_{s\in J} |x(s) - y(s)|, \end{aligned}$$

which gives a contraction in the max-norm ($\|Tx - Ty\| \leq \tau K\|x - y\|$) with the constant $\tau K$ being less than $1$ by definition of $\tau$.

Step 4 — Conclusion. The existence of a unique fixed point for $T$ now follows from the Banach fixed point Theorem 1.4 applied to the closed ball $Y$ (which is a complete metric space). □

*Remark* 1.7. The proof provides more than existence of a solution: it gives a method for *constructing* approximate solutions that converge to the exact solution in the max-norm. For one may begin with the constant function $x_0$ (or any other function in the ball $Y$) and iterate the operator $T$ to obtain a sequence of functions $(T^n x_0)(t)$ that converge in the max-norm to the solution $x(t)$. This method is the famous **Picard iteration** for solving an ODE.

To obtain continuous dependence of the solution with respect to the initial value $x_0$, see the treatment in Zeidler's monograph [Zeidler, Section 1.6]. The general principle is to assume that the contraction mapping depends continuously on some parameter, and then prove that its fixed point also varies continuously with the parameter.

## 1.3 *Application:* Fredholm equation

Suppose a kernel $K(t, s)$ is continuous on $[a, b] \times [a, b]$, for some $a < b$, and let $c = \max |K|$.

**Proposition 1.8.** *If* $|\lambda| < 1/c(b - a)$*, then for each continuous function* $v$ *on* $[a, b]$*, the Fredholm integral equation*

$$x(t) - \lambda \int_a^b K(t, s)x(s)\, ds = v(t), \qquad t \in [a, b],$$

*has a unique continuous solution* $x$*.*

*Proof.* Define $T : C[a, b] \to C[a, b]$ by

$$Tx(t) = v(t) + \lambda \int_a^b K(t, s)x(s)\, ds.$$

(It is straightforward to check that $Tx$ is a continuous function for $t \in [a, b]$.) To solve the Fredholm equation, we seek a fixed point:

$$x = Tx.$$

Clearly $T$ is a contraction with respect to the max-norm, since if $x, y \in C[a,b]$ then

$$\max_{t\in[a,b]} |Tx(t) - Ty(t)| = |\lambda| \max_{t\in[a,b]} \Big| \int_a^b K(t,s)[x(s) - y(s)]\, ds \Big| \\ \leq |\lambda|(b-a)c \max_{s\in[a,b]} |x(s) - y(s)| \tag{1.6}$$

and the contraction constant $|\lambda|(b-a)c$ is less than $1$ by hypothesis. Thus we may apply the Banach fixed point Theorem 1.4 □

*A remark that will make more sense later in the course:* The assumption $|\lambda| < 1/c(b-a)$ is a crude way of making sure that $1/\lambda$ is not an eigenvalue of the integral operator $x \mapsto \int_a^b K(\cdot, s)x(s)\, ds$ (or more precisely, that $1/\lambda$ is not in the spectrum).

***Exercise*** 1.4. Find a Fredholm equation which does *not* have a solution. *Hint.* Seek an example with $\lambda = 1/c(b-a)$.

Fredholm equations arise in renewal processes, for example, which represent birth/death processes with age-dependent fertility and mortality [Keener, Chapter 3].

## 1.4 Contraction-power mapping principle

**Theorem 1.9.** *Assume* $T : X \to X$ *is continuous and the metric space is complete. If* $T^m$ *is a contraction for some positive integer* $m$ *then* $T$ *has a unique fixed point, and the fixed point is globally attracting.*

*Proof.* Uniqueness is clear, since any point fixed by $T$ is fixed also by $T^m$ and $T^m$ has a unique fixed point by Theorem 1.4.

To prove existence of a fixed point, write $x_*$ for the globally attracting fixed point of $T^m$, with contraction constant $\alpha$. Then for an arbitrary point $x \in X$ we have

$$T^{mn}x \to x_* \qquad \text{as } n \to \infty.$$

Next we show that

$$T^{mn+k}x \to x_* \qquad \text{as } n \to \infty,$$

whenever $k$ is a fixed positive integer. Indeed, the contraction property of $T^m$ implies that

$$d(T^{mn+k}x, T^{mn}x) \leq \alpha^n d(T^k x, x) \to 0$$

as $n \to \infty$, so that $T^{mn+k}x$ has the same limit as $T^{mn}x$ as $n \to \infty$, which is $x_*$.

It follows that every sequence of iterates converges to $x$ (meaning $T^n x \to x_*$). Hence $x_*$ is a fixed point for $T$ by (1.3), using here the continuity of $T$, and $x_*$ is globally attracting. □

***Exercise*** 1.5. Give an example of a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ which is not a contraction but for which $T^2$ is a contraction. Explain your example by describing the geometric effect of the transformation on the plane.

Can you construct an example with the additional property that $|Tx| = 2|x|$ for some vector $x \in \mathbb{R}^2, x \neq 0$ (so that $T$ is *really* not a contraction)?

## 1.5 *Application:* Volterra equation

Suppose a kernel $K(t,s)$ is continuous on the triangle $\{(t,s) : a \leq s \leq t \leq b\}$, for some $a < b$, and let $c = \max |K|$.

**Proposition 1.10.** *If* $\lambda \in \mathbb{R}$ *then for each continuous function* $v$ *on* $[a,b]$*, the Volterra integral equation*

$$x(t) - \lambda \int_a^t K(t,s)x(s)\,ds = v(t), \qquad t \in [a,b],$$

*has a unique continuous solution* $x$*.*

The difference from the Fredholm situation is that for Volterra, the upper limit of integration $t$ is variable. This difference allows us to treat all $\lambda \in \mathbb{R}$, in contrast to the restricted range of $\lambda$-values in Proposition 1.8.

*Proof.* Define $T : C[a,b] \to C[a,b]$ by

$$Tx(t) = v(t) + \lambda \int_a^t K(t,s)x(s)\,ds.$$

(One must check that $Tx$ is a continuous function for $t \in [a,b]$.) To solve the Volterra equation, we seek a fixed point:

$$x = Tx.$$

The contraction-power mapping principle (Theorem 1.9) will complete the proof that $T$ has a unique fixed point, provided we prove $T$ is continuous and $T^m$ is a contraction for some positive $m$. For that it suffices to show for each $m \geq 0$ that

$$\|T^m x - T^m y\| \leq \frac{\big(|\lambda| c(b-a)\big)^m}{m!} \|x - y\|, \tag{1.7}$$

because $m = 1$ gives continuity of $T$ and letting $m \to \infty$ shows $T^m$ is a contraction for sufficiently large $m$ (noting that the coefficient on the right side of (1.7) tends to $0$ as $m \to \infty$, and so will certainly be less than $1$).

The norm estimate (1.7) is a consequence of the following pointwise estimate, which we will prove by induction:

$$|T^m x(t) - T^m y(t)| \leq \frac{\big(|\lambda| c(t-a)\big)^m}{m!} \max_{s \in [a,b]} |x(s) - y(s)|, \quad t \in [a,b]. \tag{1.8}$$

(Notice the "$t$" on the right side.) The induction basis $m = 0$ is obvious. Suppose (1.8) holds for some $m \geq 0$. Then by definition of $T$ and the induction hypothesis we have

$$\begin{aligned} |T^{m+1}x(t) - T^{m+1}y(t)| &= \Big|\lambda \int_a^t K(t,s)[T^m x(s) - T^m y(s)]\, ds\Big| \\ &\leq |\lambda| c\, \frac{\big(|\lambda| c\big)^m}{m!} \int_a^t (s-a)^m\, ds \max_{s \in [a,b]} |x(s) - y(s)| \\ &= \frac{\big(|\lambda| c\big)^{m+1}}{m!} \frac{(t-a)^{m+1}}{m+1} \max_{s \in [a,b]} |x(s) - y(s)| \end{aligned}$$

by evaluating the integral. The last formula proves (1.8) for $m + 1$, and so completes the proof. □

## 1.6 Compact operators, with application to integral operators

The goal in the next section will be to prove existence of a fixed point assuming *compactness* of the mapping, rather than contractivity. Contractivity is a rather strong property, and compactness is often easier to prove in practice. For that reason, in this section we study compact operators.

Let $X$ and $Y$ be normed linear spaces, and $N$ be a positive integer.

**Definition 1.11.** An operator $T : X \to Y$ is **compact** if it is continuous and for every bounded sequence $\{x_n\}$ in $X$, the image sequence $\{Tx_n\}$ in $Y$ has a convergent subsequence. In other words, the image of every bounded set is relatively compact (or pre-compact).

The definition extends to the case where $T$ is defined only on a subset $M \subset X$; in that case the operator must be continuous on $M$ and for every bounded sequence $\{x_n\}$ in $M$, the image sequence $\{Tx_n\}$ in $Y$ must have a convergent subsequence..

**Alert:** we do not assume the operator $T$ is linear! So please do *not* assume in this section that continuity of $T$ is equivalent to boundedness.

The next two exercises illuminate the relation between continuity and compactness.

***Exercise*** 1.6 (Finite dimensions). Show that if $T : \mathbb{R}^N \to \mathbb{R}^N$ is continuous then it is compact.

***Exercise*** 1.7 (Infinite dimensions).
(a) Show that the identity operator $I : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ is continuous but not compact.
(b) Let $T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ be the projection operator onto the first $N$ components:

$$T\big((a_1, a_2, \dots, a_N, a_{N+1}, \dots)\big) = (a_1, a_2, \dots, a_N, 0, \dots).$$

Show that $T$ is continuous and compact.

Now as an application we study nonlinear integral operators acting on continuous functions.

**Theorem 1.12** (Compactness of integral operators)**.** *Let* $E \subset \mathbb{R}^N$ *be a compact set of positive Lebesgue measure. If* $F : E \times E \times \mathbb{R} \to \mathbb{R}$ *is continuous, then the operator* $A : C(E) \to C(E)$ *defined by*

$$Au(x) = \int_E F(x, y, u(y))\, dy, \qquad x \in E,$$

*is compact.*

In the special case $F(x, y, z) = K(x, y)z$ we have $Au(x) = \int_E K(x, y)u(y)\,dy$, which means $A$ is a *linear* integral operator with kernel $K$.

*Proof.* The function $(x, y) \mapsto F(x, y, u(y))$ is continuous on the compact set $E \times E$, and thus is uniformly continuous. Hence it is straightforward to show that $Au(x)$ is a continuous function of $x$.

Assume $\{u_n\}$ is a bounded sequence in $C(E)$, say with $\|u_n\| \leq r$ for all $n$, for some $r > 0$. Let $Q = E \times E \times [-r, r]$ and $c = \max_Q |F|$. Then for all $n$ and all $x$,

$$|Au_n(x)| = \Big| \int_E F(x, y, u_n(y))\,dy \Big| \leq |E|c$$

where $|E|$ denotes the Lebesgue measure of $E$.

Thus the functions $\{Au_n\}$ are uniformly bounded. They are also uniformly equicontinuous, as follows. Given $\varepsilon > 0$, the uniform continuity of $F$ on the compact set $Q$ implies the existence of a $\delta > 0$ such that

$$|F(x, y, z) - F(x', y, z)| < \varepsilon/|E|$$

whenever $x, x', y \in E, z \in [-r, r]$ with $|x - x'| < \delta$. Hence

$$|Au_n(x) - Au_n(x')| = \Big| \int_E [F(x, y, u_n(y)) - F(x', y, u_n(y))]\,dy \Big| < |E| \cdot \varepsilon/|E| = \varepsilon$$

whenever $|x - x'| < \delta$, which gives uniform equicontinuity of the family $\{Au_n\}$ since $\delta$ is independent of $x, x'$ and $n$.

The Arzelà–Ascoli Theorem now implies that some subsequence of $\{Au_n\}$ converges uniformly (that is, in the max-norm) to a continuous function on $E$. Thus the operator $A$ maps each bounded sequence to a sequence with a convergent subsequence.

Lastly we show $A$ is a continuous mapping, by an argument similar to the previous part of the proof. Suppose $u_n \to u$ in $C(E)$, so that $\{u_n\}$ is a bounded sequence. Since $\|u_n\| \leq r$ for all $n$, we also have $\|u\| \leq r$. Given $\varepsilon > 0$, the uniform continuity of $F$ on the compact set $Q$ implies the existence of a $\delta > 0$ such that

$$|F(x, y, z) - F(x, y, z')| < \varepsilon/|E|$$

whenever $x, y \in E, z, z' \in [-r, r]$ with $|z - z'| < \delta$. Take $n$ large enough that $\|u_n - u\| < \delta$. Then

$$|Au_n(x) - Au(x)| = \Big| \int_E [F(x, y, u_n(y)) - F(x, y, u(y))]\,dy \Big| < |E| \cdot \varepsilon/|E| = \varepsilon,$$

so that $Au_n \to Au$ in $C(E)$, which is the desired continuity of $A$. □

One should think of compact operators as having finite-dimensional range, or rather as being the limits of such operators, as the next result shows.

**Proposition 1.13** (Compact operator = limit of finite rank operators)**.** *Assume $X$ and $Y$ are Banach spaces, $M$ is a bounded nonempty subset of $X$, and $T : M \to Y$.*

*If $T$ is compact then a sequence of continuous mappings $T_n : M \to Y$ exists such that for each $n$:*

- *the range $T_n(M)$ lies in a finite dimensional subspace of $Y$, and*
- *$T_n(M)$ is contained in the convex hull of $T(M)$, and*
- *$T_n$ approximates $T$ uniformly on $M$, with $\sup_{x \in M} \|T_n x - Tx\|_Y \le \frac{1}{n}$.*

(The **convex hull** of a set is the collection of all its convex combinations; see Exercise 1.12 below.)

*Proof.*
Step 1 — $\overline{T(M)}$ is compact. Consider a sequence $y_m \in \overline{T(M)}$. Each point $y_m$ can be approximated by some point $Tx_m$, with $x_m \in M$, such that $\|Tx_m - y_m\| < 1/m$. The sequence $\{x_m\}$ is bounded because $M$ is bounded, and so compactness of $T$ implies that some subsequence of $Tx_m$ converges in $Y$. The corresponding subsequence of $y_m$ must also converge, to the same limit. Thus $\overline{T(M)}$ is compact.

Step 2 — Definition of $T_n$. Fix $n \ge 1$. Since the set $\overline{T(M)}$ is covered by the open balls of radius $1/2n$ centered at points of $T(M)$, it must be covered by finitely many such balls, in view of the compactness we established in Step 1. Denote the centers of these balls by $y_j \in T(M), j = 1, \ldots, k$. Each point of $\overline{T(M)}$ then lies within distance $1/2n$ of one of the points $y_1, \ldots, y_k$.

Define nonnegative, continuous functions

$$b_j(x) = \max\left\{\frac{1}{n} - \|Tx - y_j\|_Y \, , \, 0\right\}, \qquad x \in M,$$

for $j = 1, \ldots, k$. For each $x$ our construction guarantees $\|Tx - y_l\|_Y < 1/2n$ for some $l$, so that $b_l(x) > 1/2n > 0$ for that $l$-value. Thus we may define

$$a_j(x) = \frac{b_j(x)}{\sum_{l=1}^{k} b_l(x)} \ge 0$$

and

$$T_n x = \sum_{j=1}^{k} a_j(x) y_j. \tag{1.9}$$

Step 3 — Properties of $T_n$. Obviously the range of $T_n$ lies in the span of finitely many vectors $y_1, \ldots, y_k$. Further, the definition (1.9) expresses $T_n x$ as a convex combination of points $y_j \in T(M)$, and so $T_n(M)$ lies in the convex hull of $T(M)$.

Lastly, the functions $a_j(x)$ form a partition of unity ($\sum_{j=1}^{k} a_j(x) = 1$ for each $x$), and so

$$\begin{aligned} \|T_n x - Tx\|_Y &= \|\sum_{j=1}^{k} a_j(x)(y_j - Tx)\|_Y \\ &\leq \sum_{j=1}^{k} a_j(x)\|y_j - Tx\|_Y \\ &\leq \sum_{j=1}^{k} a_j(x)\frac{1}{n} = \frac{1}{n} \end{aligned}$$

where in the final inequality we use for each $j$ that if $a_j(x) > 0$ then $b_j(x) > 0$ and so $\|y_j - Tx\|_Y < \frac{1}{n}$. □

***Exercise*** 1.8. Prove the converse of Proposition 1.13: if $T$ is approximated uniformly by the finite rank, continuous operators $T_n$, then $T$ is compact.

**Exercises on compactness and contractions**

***Exercise*** 1.9. Contractivity and compactness are non-comparable properties:
(a) Give an example of a linear operator $T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ that is a contraction but not compact.
(b) Give an example of a linear operator $T : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N})$ that is compact but not a contraction.

***Exercise*** 1.10.
(a) Give an example of a linear operator $T : C[0, 1] \to C[0, 1]$ that is a contraction but not compact.
(b) Give an example of a linear operator $T : C[0, 1] \to C[0, 1]$ that is compact but not a contraction.

**Exercises on convexity**

**Definition 1.14.** A subset $S$ of a linear space $X$ is **convex** if the line segment between any pair of points in $S$ lies within the set:

$$tx + (1-t)y \in S \qquad \text{whenever } x, y \in S \text{ and } t \in [0,1].$$

For example, in a normed linear space every open ball is convex, as follows. Suppose $x$ and $y$ belong to the open ball centered at $z$ with radius $r > 0$. Then whenever $t \in [0,1]$ we have

$$\begin{aligned} \|tx + (1-t)y - z\|_X &= \|t(x-z) + (1-t)(y-z)\|_X \\ &\leq t\|x-z\|_X + (1-t)\|y-z\|_X \\ &< tr + (1-t)r = r, \end{aligned}$$

and so $tx+(1-t)y$ belongs to the open ball. The argument is almost identical for closed balls, of course.

**Definition 1.15.** Define the **convex hull** to consist of all convex combinations of points in $S$:

$$\operatorname{conv} S = \{\sum_1^n a_j x_j : n \in \mathbb{N}, a_j \geq 0, x_j \in S, \sum_1^n a_j = 1\}.$$

***Exercise*** 1.11. Prove that if $K$ is a convex subset of $X$ then

$$K = \operatorname{conv} K.$$

That is, $K$ equals the set of its convex combinations.

***Exercise*** 1.12. Let $S \subset X$. Prove that $\operatorname{conv} S$ is the smallest convex set containing $S$.

***Exercise*** 1.13. (Simple Carathéodory theorem) Show that if $S \subset \mathbb{R}^N$ then

$$\operatorname{conv} S = \{\sum_1^{N+1} a_j x_j : a_j \geq 0, x_j \in S, \sum_1^{N+1} a_j = 1\}.$$

That is, one need only use convex combinations with $N+1$ (or fewer) terms, when dealing with sets in $\mathbb{R}^N$.

## 1.7 Brouwer and Schauder fixed point theorems

Brouwer's fixed point theorem treats *finite* dimensional spaces.

**Theorem 1.16** (Brouwer fixed point theorem)**.** *Let* $M$ *be a closed, bounded, convex, nonempty subset of a finite dimensional, normed linear space* $X$*. If* $T : M \to M$ *is continuous then* $T$ *has at least one fixed point.*

*Proof.* This seemingly innocuous result is not easy to establish. We assume that you have seen a proof elsewhere, perhaps in a course on algebraic topology (where Brouwer's Theorem is equivalent to the No-Retraction Theorem), or else you are prepared to take the theorem on trust. For more information, see [Zeidler, Chapters 12, 13]. □

For example, if a rubber ball is deformed so that the final shape is contained within the original ball, then at least one point in the ball (either inside the ball or on its boundary) must remain in its original position.

*Aside.* Brouwer's theorem says nothing about how to find the fixed point. Also, the fixed point need not be attracting; indeed, it might be repelling, as happens for the fixed point at the origin of the mapping $Tx = \sin(\pi x/2)$ on $M = [-1, 1]$.

Schauder extended Brouwer's result to *infinite* dimensions. Continuity of the mapping is no longer enough (see Exercise 1.14 at the end of the section), and so we strengthen that hypothesis to compactness.

**Theorem 1.17** (Schauder fixed point theorem)**.** *Let* $M$ *be a closed, bounded, convex, nonempty subset of a Banach space* $X$*. If* $T : M \to M$ *is compact then* $T$ *has at least one fixed point.*

*Proof.* The strategy of the proof is to approximate $T$ with finite rank operators, then apply Brouwer's fixed point theorem to obtain fixed points of the approximating operators, and finally show that these fixed points converge to a fixed point of $T$ as the approximation gets better and better.

Step 1 — Approximation with a finite rank operator. For each $n \geq 1$, there exists a finite dimensional subspace $X_n$ of $X$ and a continuous mapping $T_n : M \to M \cap X_n$ such that

$$\|T_n x - Tx\|_X \leq \frac{1}{n}, \qquad x \in M,$$

where we have used Proposition 1.13 and compactness of the operator $T$.

Step 2 — Fixed point of the finite rank operator. Let $M_n = M \cap X_n$, so that $M_n$ is a closed, bounded, convex, nonempty subset of the finite dimensional, normed linear space $X_n$. Since the restriction of $T_n : M_n \to M_n$ is continuous, Brouwer's Theorem 1.16 guarantees that $T_n$ has a fixed point $x_n \in M_n$, with $T_n x_n = x_n$.

Step 3 — Convergence of the fixed points. The sequence $\{x_n\}$ of fixed points lies in $M$, and hence is bounded. Compactness of $T$ implies that some subsequence of $\{Tx_n\}$ converges, say $Tx_{n_k} \to x \in X$, and in fact $x \in M$ because $M$ is closed and $Tx_n \in M$ for all $n$. Hence

$$\begin{aligned}
\|x_{n_k} - x\|_X &= \|T_{n_k} x_{n_k} - x\|_X \qquad \text{since } x_n \text{ is a fixed point of } T_n \\
&\le \|T_{n_k} x_{n_k} - T x_{n_k}\|_X + \|T x_{n_k} - x\|_X \\
&\le \frac{1}{n_k} + o(1) \qquad \text{by construction of } T_n \text{ and } x \\
&\to 0
\end{aligned}$$

as $k \to \infty$. Therefore $x_{n_k} \to x$, and so continuity of $T$ gives that

$$Tx = \lim_{k \to \infty} T x_{n_k} = x,$$

so that $x$ is a fixed point for $T$. □

***Exercise*** 1.14. Let $M$ be the closed unit ball in $\ell^2(\mathbb{N})$. Write $a = (a_1, a_2, a_3, \dots)$ for a typical element of $M$, with norm $\|a\| = (a_1^2 + a_2^2 + a_3^2 + \cdots)^{1/2}$, and define

$$Ta = \left(\sqrt{1 - \|a\|^2}, a_1, a_2, a_3, \dots\right).$$

Show that $T$ maps the closed unit ball continuously into the unit sphere in $\ell^2(\mathbb{N})$, that $T$ is not compact, and that $T$ has no fixed points.

Thus the Schauder fixed point theorem fails without its compactness hypothesis. Obviously this counterexample cannot be employed against the Brouwer theorem in finite dimensions, because the example relies on "adding one more dimension".

## 1.8 *Application:* Peano's existence theorem for ODEs

Consider again the first order ODE initial value problem

$$x'(t) = f\big(t, x(t)\big), \qquad x(t_0) = x_0, \tag{1.10}$$

where $t_0, x_0 \in \mathbb{R}$ are given. Assume $f$ is a continuous function on the closed rectangle

$$R = \{(t,x) : |t - t_0| \le a, |x - x_0| \le b\}$$

for some $a, b > 0$. Write $c = \max|f|$, and let $\tau < \min\{a, b/c\}$.

**Theorem 1.18** (Peano)**.** *The initial value problem* (1.10) *has at least one solution* $x(\cdot)$ *on the interval* $J = [t_0 - \tau, t_0 + \tau]$.

Peano's hypotheses are weaker than Picard's in Theorem 1.6, because here we do not assume $f$ to be Lipschitz continuous in the $x$-variable. Peano's conclusions are weaker too, because Theorem 1.18 provides only existence and not uniqueness.

*Proof.* Step 1 — Main points. First we reformulate the initial value problem as a fixed point problem for the integral operator $T : M \to M$ defined by

$$Tx(t) = x_0 + \int_{t_0}^{t} f(s, x(s))\, ds,$$

where $M$ is the closed ball in $C(J)$ centered at the constant function $x_0$ and having radius $c\tau$, just as in the proof of Picard's Theorem 1.6. Note $M$ is a convex set (why?).

Then instead of trying to show that the integral operator $T$ is a contraction, we prove it is compact by adapting the proof of Theorem 1.12 (see Step 2 below). Lastly, Schauder's Theorem 1.17 provides a fixed point $x(\cdot)$ satisfying $x = Tx$.

Step 2 — Proof of compactness for $T$. Obviously $T(M)$ is already bounded in the uniform norm, since it is a subset of $M$, which is a ball. To prove the collection of functions in $T(M)$ is uniformly equicontinuous, we let $\varepsilon > 0$ and take $\delta = \varepsilon/c$. Then whenever $x \in M$,

$$|Tx(t_1) - Tx(t_2)| = \Big|\int_{t_1}^{t_2} f(s, x(s))\, ds\Big| \le |t_1 - t_2| c < \delta c = \varepsilon$$

for all $t_1, t_2 \in J$ with $|t_1 - t_2| < \delta$. Thus the Arzelà–Ascoli Compactness Theorem implies that for every sequence of functions $x_n \in M$, some subsequence of $\{Tx_n\}$ converges uniformly (that is, in the max-norm). The limiting function belongs to $M$, since $M$ is closed.

We must still show $T$ is a continuous mapping. Given $\varepsilon > 0$, the uniform continuity of $f$ on the compact set $R$ implies the existence of $\delta > 0$ such that

$$|f(s, z_1) - f(s, z_2)| \leq \varepsilon/\tau$$

whenever $s \in J, z_1, z_2 \in [x_0 - b, x_0 + b]$ with $|z_1 - z_2| \leq \delta$. Thus if $x_1, x_2 \in M$ with $\|x_1 - x_2\| \leq \delta$, we have

$$|Tx_1(t) - Tx_2(t)| = \Big| \int_{t_0}^{t} [f(s, x_1(s)) - f(s, x_2(s))] \, ds \Big| \leq \tau \cdot \varepsilon/\tau = \varepsilon$$

for all $t \in J$, and so $\|Tx_1 - Tx_2\| \leq \varepsilon$. Hence $T$ is continuous on M. □

***Exercise*** 1.15.
(a) Let $p \geq 1$ and find a solution of the initial value problem

$$x'(t) = |x(t)|^p, \qquad x(0) = 0, \qquad t \in \mathbb{R}.$$

(b) Prove that your solution is unique on some neighborhood of the origin.
(c) Show that your solution is *not* unique if $0 < p < 1$.
(d) Discuss parts (b) and (c) in relation to the Picard and Peano theorems.

***Exercise*** 1.16 (Regularity for first order ODE). Assume $x(t)$ is a differentiable solution of

$$x'(t) = f\big(t, x(t)\big), \qquad t \in (a, b).$$

Show that if $f$ is smooth (meaning $f \in C^\infty$ as a function of two variables) then $x$ is smooth. In other words, if the data is smooth then so is the solution.

***Exercise*** 1.17 (ODE in Banach space). For this exercise, $x(t)$ is a continuous function taking values in a Banach space $X$, and $t$ is a real variable. The norm of the Banach-space valued integral is bounded by the real-valued integral of the norm:

$$\left\| \int_a^b x(t) \, dt \right\|_X \leq \int_a^b \|x(t)\|_X \, dt.$$

The fundamental theorem of calculus holds just as in the real-valued case, with the analogous proof.

The exercise is to state and prove a version of Picard's Theorem 1.6 for Banach space-valued functions $\mathsf{x}(\mathsf{t})$.

*Aside.* Peano's Theorem 1.18 fails, if the Banach space is infinite dimensional.

### Assorted exercises

***Exercise*** 1.18 (Extending a densely defined operator). Assume $\mathsf{T} : \mathcal{D} \to \mathsf{Y}$ is a bounded linear operator, where $\mathsf{Y}$ is a Banach space and $\mathcal{D}$ is a dense subspace of a Banach space $\mathsf{X}$. Prove that $\mathsf{T}$ extends uniquely to a bounded linear operator on $\mathsf{X}$.

More precisely, show how to define a bounded linear operator $\widetilde{\mathsf{T}} : \mathsf{X} \to \mathsf{Y}$ such that $\widetilde{\mathsf{T}} = \mathsf{T}$ on $\mathcal{D}$, and further show that only one such operator exists.

# Chapter 2

# Hilbert spaces and weak compactness

**References** [Folland, Chapter 5]

The next chapters of the course treat linear PDEs by Hilbert space methods, and so in this chapter we provide a foundation in Hilbert space theory. The previous chapter treated nonlinear ODEs by fixed point methods, and you might wonder whether there is a connection. There is indeed, and we will see it at the end of this course: many nonlinear PDEs can be solved by fixed point arguments, building on the linear foundation provided in this course.

## 2.1 Hilbert space basics

We begin by defining the inner product, and then we construct an example that motivates the later development of Sobolev spaces. We also prove the orthogonal decomposition theorem, and the Riesz Representation theorem for bounded linear functionals.

### Definitions, and examples on function spaces

**Definition 2.1.** Let $\mathsf{H}$ be a complex linear space. (If $\mathsf{H}$ is a real linear space, then simply replace $\mathbb{C}$ by $\mathbb{R}$ in what follows, and ignore the complex conjugates.)

An **inner product** on $\mathsf{H}$ is a function $\langle \cdot, \cdot \rangle : \mathsf{H} \times \mathsf{H} \to \mathbb{C}$ that is

- linear in the first variable: $\langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle$ for all $a, b \in \mathbb{C}, x, y, z \in H$,
- conjugate symmetric: $\langle x, y\rangle = \overline{\langle y, x\rangle}$ for all $x, y \in H$,
- positive: $\langle x, x\rangle \in (0, \infty)$ for all $x \neq 0$.

$H$ with the inner product is called an **inner product space** or **pre-Hilbert space**.

By linearity, the zero vector has vanishing inner products: $\langle 0, x\rangle = 0 \in \mathbb{C}$ since $\langle 0, x\rangle + \langle 0, x\rangle = \langle 0 + 0, x\rangle = \langle 0, x\rangle$. In particular, $\langle 0, 0\rangle = 0$.

Note the inner product is conjugate linear in the second variable, by symmetry: $\langle z, ax + by\rangle = \overline{a}\langle z, x\rangle + \overline{b}\langle z, y\rangle$.

*Example* 2.2. $\mathbb{C}^N$ has inner product $\langle x, y\rangle = x \cdot \overline{y} = \sum_{j=1}^{N} x_j \overline{y_j}$, while $\mathbb{R}^N$ has inner product $\langle x, y\rangle = x \cdot y = \sum_{j=1}^{N} x_j y_j$

**Definition 2.3.** Let

$$\|x\| = \sqrt{\langle x, x\rangle}, \qquad x \in H.$$

We prove below that $\|x\|$ is a norm on the pre-Hilbert space.

**Lemma 2.4** (Schwarz inequality)**.** $|\langle x, y\rangle| \leq \|x\|\|y\|$ *for all* $x, y \in H$*, and equality holds if and only if* $x$ *and* $y$ *are linearly dependent.*

*Proof.* If $x = 0$ then $\|x\| = 0$, in which case the result is obvious because both sides equal zero.

Suppose $x \neq 0$. Then $\|x\| > 0$, and so we can let $z = x/\|x\|$. Notice $z$ is a unit vector: $\langle z, z\rangle = 1$. Decompose $y$ into its component in the $z$-direction and then the rest: $y = u + v$ where

$$u = \langle y, z\rangle z, \qquad v = y - \langle y, z\rangle z.$$

Then $\langle u, v\rangle = 0$ by linearity and conjugate symmetry, and so

$$\|y\|^2 = \langle u + v, u + v\rangle = \langle u, u\rangle + \langle v, v\rangle \geq \langle u, u\rangle = |\langle y, z\rangle|^2,$$

which gives the Schwarz inequality after we substitute the definition of $z$. □

**Proposition 2.5** (Inner product gives a norm)**.** $\|x\|$ *gives a norm on* $H$.

*Proof.* You can check that $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{C}$ and that $\|x\| = 0$ if and only if $x = 0$.

For the triangle inequality, use linearity and the Schwarz inequality to prove that

$$\begin{aligned}\|x+y\|^2 = \langle x+y, x+y\rangle &= \|x\|^2 + 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 \\ &= \big(\|x\| + \|y\|\big)^2.\end{aligned}$$

□

**Definition 2.6.** A **Hilbert space** is a pre-Hilbert space that is complete with respect to its norm topology.

*Example* 2.7. The most important example is the Hilbert space of square integrable, complex-valued functions

$$L^2(\mu) = \Big\{f \text{ measurable} : \int_E |f|^2\, d\mu < \infty\Big\},$$

where $\mu$ is a measure on some space $E$. The inner product and norm are

$$\langle f, g\rangle_{L^2} = \int_E f\overline{g}\, d\mu \qquad \text{and} \qquad \|f\|_{L^2} = \Big(\int_E |f|^2\, d\mu\Big)^{1/2}.$$

The Schwarz inequality says

$$\Big|\int_E f\overline{g}\, d\mu\Big| \leq \|f\|_{L^2}\|g\|_{L^2}.$$

We assume you know the space $L^2$, and that it is a Hilbert space.

*Example* 2.8. Every subspace of a Hilbert space is obviously a pre-Hilbert space, using the same inner product. The subspace is a Hilbert space (complete) if and only if the subspace is closed, because in that case it contains all its limit points.

*Example* 2.9. The collection of continuous functions $C[a, b]$ is a subspace of $L^2[a, b]$ and hence forms a pre-Hilbert space under the $L^2$-inner product. It is not a Hilbert space, as we show by example.

For simplicity we consider the continuous function $u_n(x) = x^{1/n}$ in $C[-1, 1]$, where $n$ is a positive odd integer. These functions converge pointwise and in $L^2[-1, 1]$ as $n \to \infty$, but the limiting function is $u(x) = \mathrm{sign}(x)$, which is not continuous. Thus the subspace of continuous functions is not closed in $L^2$, and hence is not complete.

*Example* 2.10 (Hinting at Sobolev space theory). Now let us consider functions with derivatives. For $u \in C^1[a,b]$ (continuously differentiable functions on the closed interval) we define an inner product

$$\langle u, v\rangle_{W^{1,2}} = \int_a^b (u\bar{v} + u'\overline{v'})\, dx,$$

so that the norm-square is $\|v\|_{W^{1,2}}^2 = \|v\|_{L^2}^2 + \|v'\|_{L^2}^2$.

Then $C^1$ is a pre-Hilbert space but not a Hilbert space, with this inner product.

*Proof.* Take the interval to be $[-1,1]$ for simplicity. Consider the even function $v_n(x) = x^{1+1/n} = x \cdot x^{1/n}$ where $n$ is a positive odd integer. Then $v_n'(x) = (1+1/n)x^{1/n}$. Clearly $v_n(x) \to |x|$ and $v_n'(x) \to \text{sign}(x) = |x|'$ pointwise and in $L^2[-1,1]$, as $n \to \infty$, and so

$$\|v_n(x) - |x|\|_{W^{1,2}} \to 0$$

as $n \to \infty$. The limiting function $|x|$ (the absolute value function) is not continuously differentiable, and so $C^1$ is not complete under the $W^{1,2}$-inner product.

Incidentally, in a later chapter, we will "complete" $C^1$ to arrive at the Sobolev space $W^{1,2}$ of "weakly" differentiable functions having $1$ derivative in $L^2$. The "$W$" here stands for "weak". □

***Exercise*** 2.1. Consider the space

$$C_0^1[a,b] = \{u \in C^1[a,b] : u(a) = 0, u(b) = 0\}$$

of continuously differentiable functions that vanish at the endpoints of the interval. Obviously $C_0^1$ is a pre-Hilbert space under the $W^{1,2}$-inner product.

Show that $C_0^1$ is not a Hilbert space. Illustrate your solution with relevant graphs.

*Note.* You may choose the interval to be $[-1,1]$, to make the solution simpler.

**Definition 2.11.** Two inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$ on a pre-Hilbert space $H$ are **equivalent** if their norms are comparable, that is, if

$$\frac{1}{C}\|x\|_1 \le \|x\|_2 \le C\|x\|_1, \qquad x \in H,$$

for some $C > 0$.

***Exercise*** 2.2. Suppose you have equivalent inner products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$, that $\{x_n\}$ is a sequence in $H$, and that $x \in H$.

(i) Show that $x_n \to x$ in the $\|\cdot\|_1$-norm if and only if $x_n \to x$ in the $\|\cdot\|_2$-norm.

(ii) Deduce that $H$ is complete (a Hilbert space) with respect to one inner product if and only if it is complete with respect to the other inner product.

*Comment.* From a topological perspective, the point is that the two norms generate the same open sets, since an open ball in one norm contains an open ball in the other norm having the same center:

$$\{x : \|x\|_1 < r/C\} \subset \{x : \|x\|_2 < r\} \subset \{x : \|x\|_1 < Cr\} \qquad \forall r > 0.$$

***Exercise*** 2.3. Consider again the pre-Hilbert space

$$C_0^1[a,b] = \{u \in C^1[a,b] : u(a) = 0, u(b) = 0\}$$

of continuously differentiable functions that vanish at the endpoints of the interval, with the $W^{1,2}$-inner product

$$\langle u, v \rangle_{W^{1,2}} = \int_a^b (u\overline{v} + u'\overline{v'})\, dx.$$

Show that an equivalent inner product on $C_0^1$ is provided by the $L^2$-inner product of the derivatives:

$$\langle u, v \rangle_{\text{equiv}} = \langle u', v' \rangle_{L^2}.$$

*Hint.* One direction of the equivalence is easy. For the other direction, try integrating by parts.

***Exercise*** 2.4. Consider a domain $U \subset \mathbb{R}^N, N \geq 2$, with smooth boundary. so that Green's theorem and integration by parts are valid. Define the $W^{1,2}$-inner product for functions on this domain by

$$\langle u, v \rangle_{W^{1,2}} = \int_U (u\overline{v} + \nabla u \cdot \overline{\nabla v})\, dx, \qquad u, v \in C^1(\overline{U}).$$

Clearly $C^1(\overline{U})$ is a pre-Hilbert space with this inner product. Let

$$C_0^1(\overline{U}) = \{u \in C^1(\overline{U}) : u = 0 \text{ on } \partial U\}.$$

Show that an equivalent inner product on $C_0^1$ is provided by the $L^2$-inner product of the derivatives:

$$\langle u, v \rangle_{\text{equiv}} = \langle \nabla u, \nabla v \rangle_{L^2}.$$

*Aside.* Our treatment of Poisson's equation later in the chapter relies on this equivalent inner product.

## Othogonal decompositions

Now we gather a few useful properties of inner products, as we build up toward orthogonal decompositions.

**Lemma 2.12** (Continuity of the norm and inner product)**.** *If* $x_n \to x$ *and* $y_n \to y$*, then* $\|x_n\| \to \|x\|$ *and* $\langle x_n, y_n \rangle \to \langle x, y \rangle$*.*

*Proof.* By the triangle inequality, we have

$$\big| \|x_n\| - \|x\| \big| \leq \|x_n - x\| \to 0$$

and so $\|x_n\| \to \|x\|$. Hence by the Schwarz inequality,

$$\begin{aligned} \big| \langle x_n, y_n \rangle - \langle x, y \rangle \big| &= \big| \langle x_n, y_n - y \rangle + \langle x_n - x, y \rangle \big| \\ &\leq \|x_n\| \|y_n - y\| + \|x_n - x\| \|y\| \\ &\to 0. \end{aligned}$$

□

Write $x \perp y$ to mean that $x$ and $y$ are **orthogonal**, that is, $\langle x, y \rangle = 0$.

**Proposition 2.13** (Pythagorean theorem)**.** *If* $x_1, \dots, x_n \in H$ *with* $x_j \perp x_k$ *whenever* $j \neq k$*, then*

$$\Big\| \sum_{j=1}^{n} x_j \Big\|^2 = \sum_{j=1}^{n} \|x_j\|^2.$$

*Proof.* The left side equals $\sum_j \sum_k \langle x_j, x_k \rangle$, and the terms with $j \neq k$ vanish by orthogonality. □

**Lemma 2.14** (Parallelogram identity)**.**

$$\|x + y\|^2 + \|x - y\|^2 = 2\big(\|x\|^2 + \|y\|^2\big), \qquad x, y \in H.$$

*Proof.* Start with $\|x + y\|^2 = \|x\|^2 + 2\,\mathrm{Re}\langle x, y \rangle + \|y\|^2$. Replace $y$ with $-y$ and add the two formulas. The cross terms cancel, giving the right side of the lemma. □

The parallelogram identity deserves its name because the parallelogram spanned by vectors $x$ and $y$ in the plane has diagonal vectors $x + y$ (main diagonal) and $x - y$ (the other diagonal). The identity states that the sum of the squares of the diagonal lengths equals the sum of the squares of all four side lengths.

**Definition 2.15.** The **orthogonal complement** of a set $M \subset H$ is the collection of vectors orthogonal to every element of $M$:

$$M^\perp = \{z \in H : z \perp y \text{ for all } y \in M\}.$$

That is,

$$M^\perp = \{z \in H : \langle z, y\rangle = 0 \text{ for all } y \in M\}.$$

Clearly $M^\perp$ is a subspace of $H$, regardless of whether or not $M$ is a subspace, since if $a_1, a_2 \in \mathbb{C}$ and $z_1, z_2 \in M^\perp$ then $a_1 z_1 + a_2 z_2 \in M^\perp$ because $\langle a_1 z_1 + a_2 z_2, y\rangle = 0$ for all $y \in Y$.

Further, $M^\perp$ is closed, regardless of whether or not $M$ itself is closed, because if $z_n \in M^\perp$ and $z_n \to z$ then $z \in M^\perp$ because $\langle z, y\rangle = \lim\langle z_n, y\rangle = 0$ for all $y \in M$.

**Theorem 2.16** (Orthogonal decompositions)**.** *If $M$ is a closed subspace of a Hilbert space $H$ then $H = M \oplus M^\perp$. That is, each $x \in H$ can be written uniquely as $x = y + z$ where $y \in M, z \in M^\perp$.*

*Further, $y$ is the unique closest point in $M$ to $x$, and $z$ is the unique closest point in $M^\perp$ to $x$.*

For example, $\mathbb{R}^3$ can be decomposed as $\mathbb{R}^2 \oplus \mathbb{R}$, by choosing $M$ to be the $x_1x_2$-plane so that $M^\perp$ is the $x_3$-axis.

*Proof.* Step 1 — Uniqueness of decomposition. If $x = y + z$ and $x = y' + z'$ where $y, y' \in M$ and $z, z' \in M^\perp$, then by subtracting we find

$$y - y' = z' - z \in M^\perp.$$

Hence $y - y'$ is orthogonal to every element of $M$. In particular, it is orthogonal to itself, which means that $y - y' = 0$. Then also $z' - z = 0$, so that $y = y', z = z'$.

Step 2 — Existence of decomposition. Given $x \in H$, define

$$\delta = \operatorname{dist}(x, M) = \inf_{y \in M} \|x - y\|.$$

Take an infimizing sequence $\{y_n\}$ in $M$ such that $\|x - y_n\| \to \delta$. The sequence is Cauchy, because by the parallelogram identity,

$$\begin{aligned}
\|y_m - y_n\|^2 &= \|(x - y_m) - (x - y_n)\|^2 \\
&= 2\big(\|x - y_m\|^2 + \|x - y_n\|^2\big) - \|(x - y_m) + (x - y_n)\|^2 \\
&= 2\big(\|x - y_m\|^2 + \|x - y_n\|^2\big) - 4\|x - (y_m + y_n)/2\|^2 \\
&\leq 2\big(\delta^2 + o(1) + \delta^2 + o(1)\big) - 4\delta^2 \\
&= o(1)
\end{aligned}$$

as $m, n \to \infty$. We used that $M$ is a subspace to insure that $(y_m + y_n)/2 \in M$, so that this point has distance at least $\delta$ from $x$.

Since $H$ is complete, the Cauchy sequence has a limit $y = \lim y_n$. Note that $y \in M$ because $M$ is closed, and $\|x - y\| = \delta$ by continuity of the norm, so that $y$ is a closest point to $x$ in $M$.

Let $z = x - y$ (so that $x = y + z$). To show $z \in M^\perp$ we let $w \in M$ and show $\langle w, z\rangle = 0$, as follows. For all numbers $re^{i\theta} \in \mathbb{C}$ with $r \neq 0$ we have

$$\begin{aligned}
\delta^2 &\leq \|x - (y + re^{i\theta}w)\|^2 \qquad \text{by definition of } \delta \\
&= \|z\|^2 - 2\,\mathrm{Re}\, re^{-i\theta}\langle z, w\rangle + r^2\|w\|^2.
\end{aligned}$$

After recalling that $\|z\| = \delta$, we can divide by $r$ and let $r \to 0$ to find that $\mathrm{Re}\, e^{-i\theta}\langle z, w\rangle \leq 0$. This inequality holds for all $\theta$, and so $\langle z, w\rangle = 0$.

Step 3 — Unique closest point property. Suppose $y' \in M$, and let $u = y' - y \in M$. Then

$$\begin{aligned}
\|x - y'\|^2 &= \|x - y - u\|^2 \\
&= \|x - y\|^2 - 2\,\mathrm{Re}\langle z, u\rangle + \|u\|^2 \\
&= \|x - y\|^2 + \|u\|^2,
\end{aligned}$$

so that the minimum distance from $x$ to $M$ is attained if and only if $u = 0$, that is, $y = y'$.

Similarly, suppose $z' \in M^\perp$, and let $v = z' - z \in M^\perp$. Then

$$\begin{aligned}
\|x - z'\|^2 &= \|x - z - v\|^2 \\
&= \|x - z\|^2 - 2\,\mathrm{Re}\langle y, v\rangle + \|v\|^2 \\
&= \|x - z\|^2 + \|v\|^2,
\end{aligned}$$

so that the minimum distance from $x$ to $M^\perp$ is attained is attained if and only if $v = 0$, that is, $z = z'$. □

***Exercise*** 2.5. Consider the Hilbert space $L^2(\mu)$, where $\mu$ is a finite measure on some space $E$ (so that $\mu(E) < \infty$). Take $M$ to be the subspace of constant functions. Your task is to identify the orthogonal complement $M^\perp$, both in precise mathematical terms and in descriptive intuitive terms.

## Bounded linear functionals and Riesz Representation

Next we develop a result of fundamental importance in the course: the Riesz Representation theorem for bounded linear functionals. We will solve Poisson's equation with its help, later in the chapter.

**Definition 2.17.** A **functional** is a scalar-valued function $F : H \to \mathbb{C}$. Call the functional

- **linear** if $F(ax + by) = aF(x) + bF(y)$ for all $a, b \in \mathbb{C}$ and $x, y \in H$,
- **bounded** if $|F(x)| \leq C\|x\|$ for all $x \in H$, for some constant $C \geq 0$.

The **norm** $\|F\|$ of a bounded functional is the smallest such constant $C$.

Bounded linear functionals are Lipschitz continuous, since

$$|F(x) - F(y)| = |F(x - y)| \leq C\|x - y\|.$$

*Example* 2.18 (Inner products give bounded linear functionals). Fix $y \in H$ and define $F(x) = \langle x, y\rangle$ for all $x \in H$. Then $F$ is clearly a linear functional, and it is bounded with constant $C = \|y\|$, since $|F(x)| \leq \|x\|\|y\|$ by the Schwarz inequality

The next theorem says that **every** bounded linear functional on a Hilbert space arises from an inner product.

**Theorem 2.19** (Riesz Representation)**.** *If $F$ is a bounded linear functional on a Hilbert space $H$, then $F$ is given by the inner product against a unique vector $y \in H$:*

$$F(x) = \langle x, y\rangle, \qquad x \in H,$$

*and $\|F\| = \|y\|_H$.*

*Proof.* Step 1 — Uniqueness. Suppose $F$ is represented by both $y_1$ and $y_2$. Then for all $x$,

$$0 = F(x) - F(x) = \langle x, y_1 - y_2\rangle.$$

Choosing $x = y_1 - y_2$ shows that $y_1 - y_2 = 0$.

Step 2 — Existence. Let $M = \ker(F) = \{x \in H : F(x) = 0\}$. Note $M$ is closed since $F$ is continuous, and $M$ is a subspace since $F$ is linear. By the Orthogonal Decompositon Theorem 2.16, we have $H = M \oplus M^\perp$.

If $M^\perp$ contains only the zero vector, then $H = M$ and so $F(x) = 0$ for all $x$, in which case we may take $y = 0$. Thus we may assume $M^\perp$ contains a nonzero vector $z$. By rescaling, we may suppose $z \in M^\perp$ is a unit vector.

Given $x \in H$, let

$$w = F(x)z - F(z)x,$$

so that $F(w) = 0$ by linearity. That is, $w \in M$, and so $w \perp z$. Hence

$$0 = \langle w, z\rangle = F(x)\langle z, z\rangle - F(z)\langle x, z\rangle = F(x) - \langle x, \overline{F(z)}z\rangle.$$

Hence $F(x) = \langle x, y\rangle$ for all $x$, where we take $y = \overline{F(z)}z$.

Step 3 — Norm of the functional. Since $|F(x)| \leq \|x\|_H \|y\|_H$ by the Schwarz inequality, we have $\|F\| \leq \|y\|_H$. Choosing $x = y$ shows that $\|F\| = \|y\|_H$. □

The Riesz Representation Theorem says that the dual space (the space of bounded linear functionals) is isomorphic to the Hilbert space itself.

## 2.2 *Application:* Weak solution of Poisson's equation

The material in this section is not rigorous. Rather, we aim to indicate how functional analysis provides a framework for constructing solutions of elliptic PDEs, and to identify some of the challenges ahead.

Consider a bounded domain $U \subset \mathbb{R}^N$. Suppose we want to solve Poisson's equation

$$\begin{aligned} -\Delta u &= f \quad \text{in } U, \\ u &= 0 \quad \text{on } \partial U, \end{aligned}$$

where the real-valued function $f \in L^2(U)$ is given. If we multiply the equation by a real-valued function $v \in C_0^1(\overline{U})$ (so that $v$ equals zero on the boundary) and then apply Green's Theorem (integration by parts), we arrive at the "weak" form of the equation:

$$\int_U \nabla u \cdot \nabla v \, dx = \int_U f v \, dx \tag{2.1}$$

for all $v \in C^1_0(\overline{U})$.

To find a function $u$ that satisfies this weak form of the equation, we define a linear functional $F(v) = \int_U fv\,dx$. This functional is bounded on $C^1_0(\overline{U})$ with respect to the $W^{1,2}$-norm, because

$$\begin{aligned} |F(v)| &\leq \|f\|_{L^2(U)}\|v\|_{L^2(U)} \qquad \text{by the Schwarz inequality} \\ &\leq \|f\|_{L^2(U)}\|v\|_{W^{1,2}(U)} \end{aligned}$$

by definition of the $W^{1,2}$-inner product and norm. Recall the comparable norm $\|v\|_{\text{equiv}} = \|\nabla v\|_{L^2(U)}$ coming from the equivalent inner product $\langle v, w\rangle_{\text{equiv}} = \int_U \nabla v \cdot \nabla w\,dx$ that we studied in Exercise 2.4. Comparability of the norms implies boundedness of $F$ with respect to this new norm:

$$|F(v)| \leq (\text{const.})\|f\|_{L^2(U)}\|v\|_{\text{equiv}}.$$

We would like to apply the Riesz Representation Theorem to $F$, but cannot, because $C^1_0$ is not a Hilbert space — Exercise 2.1 showed that it is not complete, even in one dimension. Let us ignore this obstacle for a moment. Applying the Riesz Representation Theorem 2.19 would give a real-valued function $u$ representing the functional with respect to the equivalent inner product:

$$F(v) = \langle v, u\rangle_{\text{equiv}}, \qquad v \in C^1_0(\overline{U}).$$

This last equation is exactly Poisson's equation in its weak form (2.1), as we wanted to solve.

Much remains to be done in order to turn this sketch of a proof into a rigorous argument. First, we must "complete" $C^1_0(\overline{U})$ into a genuine Hilbert space, called $W^{1,2}_0(U)$. We will construct this Sobolev space in Chapter 3, and revisit the above solution method in Exercise 3.14. Second, we must show that a weak solution of Poisson's equation is in fact a classical, twice-differentiable solution of the equation, provided the data $f$ is smooth enough; this second task leads us into elliptic regularity theory in Chapter 5.

## 2.3 Orthonormal bases

The concept of an orthonormal basis (ONB) is central to Hilbert space theory, and so we will begin by reminding the reader of the subtleties of the concept in an infinite dimensional space. An ONB provides a simple example of a bounded sequence that has no convergent subsequence, which leads immediately to the realization that a closed bounded set (such as the unit ball) need not be compact, in an infinite dimensional space. To salvage the concept of compactness, we will weaken the topology, and thus arrive in the next section at a topic of central importance for PDEs: the weak compactness of closed balls in a Hilbert space.

In this section, we assume $H$ is a Hilbert space. For simplicity, we will consider only countable orthonormal sets.

**Definition 2.20.** A set $\{x_j\} \subset H$ is **orthonormal** if

$$\langle x_j, x_k \rangle = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

In particular, $\|x_j\| = 1$.

*Example* 2.21.

The standard unit vectors in $\mathbb{R}^N$ are orthonormal.

The standard unit vectors in $\mathbb{C}^N$ are orthonormal.

The standard unit vectors in $\ell^2(\mathbb{N})$ are orthonormal:

$$e_j = (0, \ldots, 0, 1, 0, \ldots)$$

where the $1$ is in position $j$.

The Fourier exponentials $\{u_j(t) = e^{2\pi i j t}\}_{j \in \mathbb{Z}}$ are orthonormal in $L^2[0,1]$.

The Fourier sines $\{v_k(x) = \sqrt{2/\pi}\sin(kx)\}_{k=1}^{\infty}$ are orthonormal in $L^2[0,\pi]$.

***Exercise*** 2.6. Show that the normalized sines $\{\sqrt{2/\pi(k^2+1)}\sin(kx)\}_{k=1}^{\infty}$ are orthonormal with respect to the $W^{1,2}$-inner product on $[0,\pi]$.

***Exercise*** 2.7. Let $h = 1_{[0,1/2)} - 1_{[1/2,1)}$ where "$1_E$" denotes the indicator (or characteristic) function of a set $E$. Define the **Haar wavelet functions** by

$$h_{j,k}(x) = 2^{j/2} h(2^j x - k), \qquad j, k \in \mathbb{Z}.$$

(a) Sketch the graphs of $h_{j,k}(x)$ for $j = 0$ and $k = -1, 0, 1$, and for $j = 1$ and $k = -2, -1, 0, 1$.
(b) Prove that $\{h_{j,k}\}_{j,k \in \mathbb{Z}}$ is orthonormal in $L^2(\mathbb{R})$.

The size of inner products taken against an orthonormal system are controlled by the norm.

**Proposition 2.22** (Bessel's inequality)**.** *If* $\{x_j\}$ *is orthonormal, then*

$$\sum_j |\langle x, x_j\rangle|^2 \le \|x\|^2, \qquad x \in H.$$

*Proof.* Suppose B is a finite set of j-values, so that we do not need to worry about convergence of the sums in the following proof. We will show $\sum_{j\in B}|\langle x, x_j\rangle|^2 \le \|x\|^2$. Then letting B expand to the full index set proves the proposition.

We have

$$\begin{aligned}
0 &\le \Big\|x - \sum_{j\in B}\langle x, x_j\rangle x_j\Big\|^2 \\
&= \|x\|^2 - 2\sum_{j\in B} \mathrm{Re}\,\big\langle x, \langle x, x_j\rangle x_j\big\rangle + \Big\|\sum_{j\in B}\langle x, x_j\rangle x_j\Big\|^2 \\
&= \|x\|^2 - 2\sum_{j\in B} |\langle x, x_j\rangle|^2 + \sum_{j\in B} |\langle x, x_j\rangle|^2 \qquad (2.2)\\
&= \|x\|^2 - \sum_{j\in B} |\langle x, x_j\rangle|^2, \qquad (2.3)
\end{aligned}$$

where we used orthonormality of the $x_j$ to get (2.2) (that is, we used the Pythagorean theorem 2.13). □

The fundamental operators associated with an orthonormal collection of vectors are **analysis** $A : x \mapsto \{\langle x, x_j\rangle\}$ and **synthesis** $S : \{c_j\} \to \sum_j c_j x_j$.

**Corollary 2.23** (Analysis $A : H \to \ell^2$)**.** *If* $\{x_j\}$ *is an orthonormal collection of vectors and* $x \in H$*, then the coefficient sequence*

$$c_j = \langle x, x_j\rangle$$

*belongs to* $\ell^2$ *with* $\|\{c_j\}\|_{\ell^2} \le \|x\|_H$*. Thus the analysis operator maps* H *into* $\ell^2$ *and does not increase norms.*

*Proof.* Immediate from Bessel's inequality. □

**Proposition 2.24** (Synthesis $S : \ell^2 \to H$)**.** *If* $\{x_j\}$ *is an orthonormal collection of vectors in a Hilbert space* H *and the coefficient sequence* $\{c_j\} \subset \mathbb{C}$ *belongs to* $\ell^2$*, then the series*

$$\sum_j c_j x_j$$

*converges unconditionally in* $H$. *(Unconditional convergence means that the partial sums converge to the same limit regardless of the order in which the terms of the series are put.) Further, the vector* $x = \sum_j c_j x_j$ *has* $\langle x, x_j \rangle = c_j$ *and* $\|x\|_H = \|\{c_j\}\|_{\ell^2}$. *Thus the synthesis operator maps* $\ell^2$ *into* $H$ *and preserves norms.*

***Exercise*** 2.8. Prove Proposition 2.24.

The central result of the section is:

**Theorem 2.25** (Completeness Criteria). *Let* $\{x_j\}$ *be an orthonormal set in a Hilbert space* $H$. *Then the following are equivalent:*
*(a) [Completeness] If* $x \in H$ *is orthogonal to every* $x_j$ *then* $x = 0$.
*(b) [Parseval's identity]*

$$\|x\|^2 = \sum_j |\langle x, x_j \rangle|^2, \qquad x \in H.$$

*(c) [Orthonormal expansions] Each* $x \in H$ *can be expressed as* $x = \sum_j \langle x, x_j \rangle x_j$ *where the sum converges unconditionally in* $H$ *(meaning, regardless of the order in which the terms in the series are put).*

*Proof.*
(c) $\Longrightarrow$ (b): Bessel's inequality already proves "$\geq$" in (b). To prove "$\leq$", we let $\varepsilon > 0$ and take a sufficiently large partial sum of the series in condition (c), getting that $\|x - \sum_{j=1}^J \langle x, x_j \rangle x_j\| \leq \varepsilon$ for some index $J$. Then

$$\begin{aligned} \|x\|^2 - \sum_j |\langle x, x_j \rangle|^2 &\leq \|x\|^2 - \sum_{j=1}^J |\langle x, x_j \rangle|^2 \\ &= \Big\|x - \sum_{j=1}^J \langle x, x_j \rangle x_j\Big\|^2 \quad \text{by identity (2.3) with } B = \{1, \dots, J\} \\ &\leq \varepsilon^2, \end{aligned}$$

and now condition (b) follows by letting $\varepsilon \to 0$.
(b) $\Longrightarrow$ (a): obvious
(a) $\Longrightarrow$ (c): The series $\sum_{j=1}^\infty \langle x, x_j \rangle x_j$ converges because its partial sums are Cauchy: indeed, by the Pythagorean theorem we have

$$\Big\| \sum_{j=n+1}^{n+m} \langle x, x_j \rangle x_j \Big\|^2 = \sum_{j=n+1}^{n+m} |\langle x, x_j \rangle|^2 \to 0$$

as $n \to \infty$, in view of the convergence of the series in Bessel's inequality.

Write $y = x - \sum_{j=1}^{\infty} \langle x, x_j \rangle x_j$, so that $\langle y, x_k \rangle = 0$ for all $k$ by using the orthonormality. Hence by (a) we have $y = 0$, which means $x = \sum_{j=1}^{\infty} \langle x, x_j \rangle x_j$ as we wanted to show for (c). The unconditionality of the convergence is clear, since condition (a) is independent of the ordering of the orthonormal set and so our argument shows that regardless of the ordering, the series converges to the same value $x$. □

**Definition 2.26.** An **orthonormal basis (ONB)** for $H$ is a complete orthonormal set $\{x_j\}$, meaning conditions (a), (b) and (c) hold in the last theorem.

**Proposition 2.27.** *Every Hilbert space has an ONB.*

See [Folland, Proposition 5.28] for the short proof, which uses Zorn's Lemma. Note the ONB might be uncountable.

In the next theorem, a Hilbert space $H$ is called **separable** if it contains a countable, dense subset. All the spaces we will consider are separable.

**Theorem 2.28.** *A Hilbert space* $H$ *is separable if and only if it possesses a countable ONB, and in that case every ONB for* $H$ *is countable.*

*Proof.* "$\Longleftarrow$": If $H$ has a countable ONB then every point in $H$ can be approximated arbitrarily well by finite linear combinations of the basis vectors, using only rational coefficients. Such linear combinations form a countable dense set in $H$.

"$\Longrightarrow$": This direction is a pleasant exercise using Gram–Schmidt orthonormalization of the countable dense set, combined with Theorem 2.25(a) to prove completeness of the resulting basis (see [Folland, Proposition 5.29]).

The final statement is left to the reader. □

We will rarely need the last two results, because later in the course we give a direct proof of existence for orthonormal bases consisting of eigenfunctions of the Laplacian and related differential operators, in the function spaces that we employ for studying PDEs.

***Exercise*** 2.9. Show that the standard unit vectors form an ONB for $\ell^2(\mathbb{N})$.

***Exercise*** 2.10. Show that the Fourier exponentials $\{u_j(t) = e^{2\pi i j t}\}_{j \in \mathbb{Z}}$ form an ONB for $L^2[0, 1]$. *Note.* It is best to defer this exercise until later in the course, for then we can apply the spectral theory of selfadjoint operators: the $u_j$ are the eigenfunctions of the second derivative operator $-u_j'' = 4\pi^2 j^2 u_j$.

***Exercise*** 2.11. Show that the Fourier sines $\{v_k(x) = \sqrt{2/\pi}\sin(kx)\}_{k=1}^{\infty}$ form an ONB for $L^2[0,\pi]$. *Note.* Again, it is best to defer this exercise until later in the course.

***Exercise*** 2.12. Show that the Haar functions form an ONB for $L^2(\mathbb{R})$.

***Exercise*** 2.13 (Nonorthogonal expansions).
(a) Find four vectors $v_1, v_2, v_3, v_4 \in \mathbb{R}^2$, all having the same length and pointing in different directions, such that the expansion property holds:

$$x = \sum_{j=1}^{4} (x \cdot v_j) v_j, \qquad x \in \mathbb{R}^2.$$

(b) Find three vectors $v_1, v_2, v_3 \in \mathbb{R}^2$, all having the same length and pointing in different directions, such that the expansion property holds:

$$x = \sum_{j=1}^{3} (x \cdot v_j) v_j, \qquad x \in \mathbb{R}^2.$$

This exercise shows the expansion property holds for certain systems of vectors that are not orthonormal bases. To learn more about such nonorthogonal expansions, you can read up on "frames" in harmonic analysis, especially "Parseval frames".

***Exercise*** 2.14 (Poisson equation in 1-dimension). Suppose $-u'' = f$ on the interval $[0,\pi]$, with Dirichlet boundary conditions $u(0) = 0, u(\pi) = 0$. We could solve the equation by simply integrating twice, but in this exercise we will instead solve using the $L^2$-ONB $\{v_k(x) = \sqrt{2/\pi}\sin(kx)\}_{k=1}^{\infty}$ of eigenfunctions of the second derivative operator, because this approach generalizes to higher dimensions and more sophisticated differential operators.

Write $\lambda_k = k^2$ so that $-v_k'' = \lambda_k v_k$, let $f \in L^2[0,\pi]$, and define

$$u = \sum_k \frac{1}{\lambda_k} \langle f, v_k \rangle_{L^2} v_k.$$

(i) Check that $-u'' = f$ formally (that is, without worrying about convergence of series or about the validity of differentiating through a series).
(ii) Show that the series for $u$ converges in $L^2[0,\pi]$, meaning the definition of $u$ makes sense.

(iii) Prove that $-u'' = f$ weakly, in the sense that $-\langle u, v'' \rangle_{L^2} = \langle f, v \rangle_{L^2}$ for all $v \in C^2[0, \pi]$ with $v(0) = 0, v(\pi) = 0$. (This definition of weak solution is slightly weaker than we will use in the next chapter, but it is good enough for now.)

***Exercise*** 2.15 (Diffusion equation in 1-dimension). With $v_k$ and $f$ as in the previous problem, define

$$u(x, t) = \sum_k e^{-\lambda_k t} \langle f, v_k \rangle_{L^2} v_k(x).$$

(i) Check that $u_t = u_{xx}$ formally (that is, without worrying about convergence of series or about the validity of differentiating through a series) and that $u$ has initial condition $u(\cdot, 0) = f$.
(ii) Show that the series for $u$ converges in $L^2[0, \pi]$, at each fixed time $t \geq 0$, so that the definition of $u$ makes sense.
(iii) Prove that $u_t = u_{xx}$ weakly, in the sense that $\frac{d}{dt} \langle u, v \rangle_{L^2} = \langle u, v_{xx} \rangle_{L^2}$ for all $v \in C^2[0, \pi]$ with $v(0) = 0, v(\pi) = 0$. *Hint.* As part of your proof, show that $\{\lambda_k \langle v_k, v \rangle\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$.

***Exercise*** 2.16 (Wave equation in 1-dimension). With $v_k$ as in the previous problems, and with $f, g \in L^2[0, \pi]$, define

$$u(x, t) = \sum_k \left[ \cos(\sqrt{\lambda_k} t) \langle f, v_k \rangle_{L^2} + \frac{1}{\sqrt{\lambda_k}} \sin(\sqrt{\lambda_k} t) \langle g, v_k \rangle_{L^2} \right] v_k(x).$$

(i) Check that $u_{tt} = u_{xx}$ formally (that is, without worrying about convergence of series or about the validity of differentiating through a series). Similarly check formally that $u$ has initial displacement $u(\cdot, 0) = f$ and initial velocity $u_t(\cdot, 0) = g$.
(ii) Show that the series for $u$ converges in $L^2[0, \pi]$, at each fixed time $t \in \mathbb{R}$, so that the definition of $u$ makes sense.
(iii) Prove that $u_{tt} = u_{xx}$ weakly, in the sense that $\frac{d^2}{dt^2} \langle u, v \rangle_{L^2} = \langle u, v_{xx} \rangle_{L^2}$ for all $v \in C^2[0, \pi]$ with $v(0) = 0, v(\pi) = 0$. *Hint.* As part of your proof, show that $\{\lambda_k \langle v_k, v \rangle\}_{k=1}^{\infty} \in \ell^2(\mathbb{N})$.

*Remark* 2.29. We would like to apply the Hilbert space methods of the previous exercises to more sophisticated differential operators (such as Sturm–Liouville operators), and to domains in higher dimensions. The first obstacle is that in those situations we do not have an ONB of eigenfunctions. Later in the course we will prove such ONBs exist, even though it is generally impossible to find explicit formulas for them.

Two other issues we must tackle later in the course are the Dirichlet boundary condition (in what sense does $u(x)$ approach $0$ as $x$ approaches the boundary?) and the regularity of the solution (is it smooth enough to satisfy the PDE classically rather than just weakly?).

## 2.4 Weak compactness

A set that is closed and bounded in Euclidean space is automatically compact, by the Bolzano–Weierstrass theorem. It would be nice if closed bounded sets in a Hilbert space (such as Sobolev space) were compact, because then we could solve PDEs by constructing sequences of approximate solutions and extracting convergent subsequences. Unfortunately, "closed and bounded" does not imply "compact" in infinite dimensions, as the next exercise vividly reveals.

***Exercise*** 2.17. Consider an orthonormal sequence $\{x_n\}_{n=1}^{\infty}$ in an infinite dimensional Hilbert space. (Such a sequence always exists, by Proposition 2.27. Also, we constructed several examples of such orthonormal sequences in the last section, such as the standard unit vectors in $\ell^2(\mathbb{N})$ and the Fourier exponentials in $L^2[0,1]$.)

Show that no subsequence of $\{x_n\}$ converges. Hence the closed bounded ball $\{x \in H : \|x\| \leq 1\}$ in the infinite dimensional Hilbert space $H$ is not compact.

We will rescue a kind of compactness (and with it our prospects for solving PDEs) by weakening the topology: we will seek only weak compactness, or more precisely, weak sequential compactness. First we need to define weak convergence.

**Definition 2.30.** A sequence $\{x_n\}$ in the Hilbert space $H$ is said to **converge weakly** to $x \in H$ if

$$F(x_n) \to F(x)$$

for all bounded linear functionals $F$. In that case we write that $x_n \rightharpoonup x$ weakly in $H$. Equivalently, $x_n \rightharpoonup x$ weakly in $H$ if

$$\langle x_n, y\rangle \to \langle x, y\rangle$$

for each $y \in H$. (Here we use the Riesz Representation Theorem 2.19 to express each bounded linear functional as an inner product against some vector in the space.)

This notion of weak convergence comes from a "weak topology" on the Hilbert space, but we will not need to develop this topology.

Note that $x_n \rightharpoonup x$ weakly if and only if $x_n - x \rightharpoonup 0$ weakly.

***Exercise*** 2.18 (Norm convergence implies weak convergence). Show that if $x_n \to x$ then $x_n \rightharpoonup x$ weakly.

***Exercise*** 2.19 (Uniqueness of the weak limit). Suppose $x_n \rightharpoonup x$ and $x_n \rightharpoonup y$. Show that $x = y$.

***Exercise*** 2.20 (Norm of the weak limit). Show that if $x_n \rightharpoonup x$ weakly then

$$\|x\| \leq \liminf_{n \to \infty} \|x_n\|.$$

***Exercise*** 2.21. Show that every orthonormal sequence in a Hilbert space converges weakly to $0$. (The sequences obviously do not converge in norm to $0$.)

Let us examine weak convergence in $L^2$.

***Exercise*** 2.22. Three ways in which a sequence of square-integrable functions can converge weakly to $0$.

(i) Escaping out to infinity: $u_n(t) = 1_{[n,n+1)}(t)$.

Show that $\|u_n\|_{L^2} = 1$ and $u_n \rightharpoonup 0$ weakly in $L^2(\mathbb{R})$.

(ii) Escaping up to infinity: $v_n(t) = 2^{n/2} 1_{[2^{-n}, 2^{-n+1})}$.
Show that $\|v_n\|_{L^2} = 1$ and $v_n \rightharpoonup 0$ weakly in $L^2(\mathbb{R})$.

(iii) Oscillating itself to death: $w_n(t) = \sqrt{2/\pi} \sin(nt)$ for $t \in [0, \pi]$ and $w_n(t) = 0$ otherwise.
Show that $\|w_n\|_{L^2} = 1$ and $w_n \rightharpoonup 0$ weakly in $L^2(\mathbb{R})$.

Now we return to general Hilbert spaces.

**Theorem 2.31** (Weak sequential compactness of closed balls; special case of Banach–Alaoglu). *If* $H$ *is a separable Hilbert space, then every bounded sequence* $\{x_n\}$ *contains a subsequence that converges weakly.*

*Proof.* In the finite dimensional case, boundedness implies norm convergence of a subsequence and hence weak convergence of the subsequence. So, we may assume $H$ is infinite dimensional.

Step 1 — Finding a good subsequence by compactness in $\mathbb{C}$. Let $\{u_j\}_{j=1}^{\infty}$ be a countable ONB for $H$, which exists by Theorem 2.28. Suppose $\{x_n\}_{n=1}^{\infty}$ is a sequence with norms bounded by some constant $C$, so that by the Schwarz inequality

$$|\langle x_n, u_j\rangle| \leq C, \qquad \forall n, j.$$

By Bolzano–Weierstrass we can find a subsequence of $\{x_n\}$ which we label $\{x_{n,1}\}_{n=1}^{\infty}$ such that the sequence of complex numbers $\{\langle x_{n,1}, u_1\rangle\}_{n=1}^{\infty}$ converges. By passing to successive subsequences we find subsequences of $\{x_n\}$ which we label as $\{x_{n,1}\}_{n=1}^{\infty}, \{x_{n,2}\}_{n=1}^{\infty}$ and so on, such that the $k$-th sequence $\{x_{n,k}\}_{n=1}^{\infty}$ has the property that $\{\langle x_{n,k}, u_j\rangle\}_{n=1}^{\infty}$ converges for each $j = 1, 2, \ldots, k$.

Step 2 — Constructing the weak limit. Consider the "diagonal" subsequence $\{y_n\}_{n=1}^{\infty} = \{x_{n,n}\}_{n=1}^{\infty}$. For each $j$, define a complex number

$$c_j = \lim_{n\to\infty} \langle y_n, u_j\rangle, \tag{2.4}$$

noting that this limit exists by construction in Step 1. Then $\{c_j\} \in \ell^2(\mathbb{N})$ since

$$\begin{aligned}
\sum_j |c_j|^2 &= \sum_j \lim_{n\to\infty} |\langle y_n, u_j\rangle|^2 \\
&\leq \liminf_{n\to\infty} \sum_j |\langle y_n, u_j\rangle|^2 \qquad \text{by Fatou's lemma} \\
&\leq \liminf_{n\to\infty} \|y_n\|^2 \qquad \text{by Bessel's inequality for the ONB} \\
&\leq C^2 < \infty
\end{aligned}$$

since $\{y_n\}$ is a subsequence of the bounded sequence $\{x_n\}$. Since the coefficients belong to $\ell^2$ we may define

$$y = \sum_j c_j u_j,$$

where the series converges unconditionally in $H$ by Proposition 2.24.

Step 3 — Proving weak convergence. We claim that $y_n \rightharpoonup y$ weakly in $H$. Certainly for each fixed $k$ and any vector of the form

$$z = \sum_{j=1}^{k} d_j u_j \in \operatorname{span}\{u_1, \ldots, u_k\}$$

we have by (2.4) that

$$\lim_n \langle y_n, z\rangle = \sum_{j=1}^{k} c_j \overline{d_j} = \langle y, z\rangle.$$

Thus we have proved the equality needed for weak convergence, but only for a dense class of vectors $z$. Fortunately that suffices, as the next lemma shows. □

**Lemma 2.32.** *If* $\{y_n\}$ *is a bounded sequence and* $\langle y_n, z\rangle \to \langle y, z\rangle$ *as* $n \to \infty$, *for all* $z$ *in a dense subset of* $H$, *then* $y_n \rightharpoonup y$ *weakly.*

*Proof.* Fix $w \in H$ and let $\varepsilon > 0$. Choose $z$ in the dense subset to satisfy $\|w - z\| < \varepsilon$. Note that $\langle y_n, z\rangle \to \langle y, z\rangle$ by hypothesis. Hence

$$\begin{aligned}\limsup_n |\langle y_n, w\rangle - \langle y, w\rangle| &\leq \limsup_n |\langle y_n - y, w - z\rangle| + \limsup_n |\langle y_n - y, z\rangle| \\ &\leq \limsup_n \|y_n - y\| \|w - z\| + 0 \\ &\leq (\sup_n \|y_n\| + \|y\|)\varepsilon.\end{aligned}$$

In the last line, the supremum is finite due to boundedness of the sequence $\{y_n\}$. Letting $\varepsilon \to 0$ now shows that $\limsup_n |\langle y_n, w\rangle - \langle y, w\rangle| = 0$, which gives weak convergence of $y_n$ to $y$. □

We have seen that boundedness implies weak convergence of some subsequence. In the converse direction we have:

**Proposition 2.33** (Weak convergence implies norm boundedness). *If* $x_n \rightharpoonup x$ *weakly in a Hilbert space* $H$ *then*

$$\sup_n \|x_n\| < \infty.$$

*Proof.* The proof uses the Uniform Boundedness Principle (exercise). □

# Chapter 3

# Sobolev spaces

**References** [Evans, Chapter 5]

**Notation and assumptions** Throughout this chapter, $U$ is a bounded domain in $\mathbb{R}^N, N \geq 1$. Functions in this section are assumed to be real-valued.

This chapter develops the theory of weak derivatives, with Sobolev spaces being spaces of functions having weak derivatives in $L^p$. The following chapter will apply Sobolev theory to prove existence and regularity of solutions to linear PDEs.

## 3.1 Green's theorem, and integration by parts

The **support** of a function $\phi$ is the closure of the set where it is nonzero:

$$\mathrm{supp}(\phi) = \overline{\{x \in U : \phi(x) \neq 0\}},$$

where the closure is taken relative to $U$. The space of smooth functions in $U$ with compact support is

$$C_c^\infty(U) = \{\phi \in C^\infty(U) : \mathrm{supp}(\phi) \text{ is a compact subset of } U\}.$$

Such functions equal zero on a neighborhood of the boundary. We define $C_c^k(U)$ similarly.

**Divergence Theorem**

We assume that you know about the Divergence theorem (also called Gauss's theorem):

$$\int_U \nabla \cdot F \, dx = \int_{\partial U} F \cdot n \, dS$$

where $F : \overline{U} \to \mathbb{R}^N$ is a vector field that is continuously differentiable, $\partial U$ is $C^1$-smooth, and $n$ is the outward normal vector field on $\partial U$. (These smoothness assumptions can be weakened, but are good enough for our purposes.)

In $1$ dimension, the domain $U$ is an interval, and the smoothness assumption on $\partial U$ is irrelevant. In this case the Divergence Theorem is simply the Fundamental Theorem of Calculus: $\int_{(a,b)} F' \, dx = F(b) - F(a)$, where the negative sign on $F(a)$ indicates the leftward orientation of the outward normal vector at the boundary point $x = a$, for the interval $(a, b)$.

**Integration by parts**

Let $u, v \in C^1(\overline{U})$ and choose $F = (0, \ldots, 0, uv, 0, \ldots, 0)$ with $uv$ appearing in the $j$-th position. Then an Integration by Parts formula follows from the Divergence theorem :

$$\int_U u_{x_j} v \, dx = -\int_U u v_{x_j} \, dx + \int_{\partial U} u v n_j \, dS$$

where $n_j$ is the $j$th component of the normal vector.

If $u$ or $v$ belongs to $C_c^1(U)$ then the boundary terms vanish and so

$$\int_U u_{x_j} v \, dx = -\int_U u v_{x_j} \, dx.$$

This last formula holds without assuming the boundary to have any smoothness, because the compact support hypothesis means we need only integrate over a domain slightly smaller than $U$, and the smaller domain can be chosen to have smooth boundary.

**Green's formulas**

Next, choosing $F = u \nabla v$ in the Divergence theorem yields Green's First Formula:

$$\int_U (\nabla u \cdot \nabla v + u \Delta v) \, dx = \int_{\partial U} u \frac{\partial v}{\partial n} \, dS \tag{3.1}$$

for $u \in C^1(\overline{U}), v \in C^2(\overline{U})$. Interchanging the roles of $u$ and $v$ and then subtracting the two formulas gives Green's Second Formula:

$$\int_U (u\Delta v - v\Delta u)\,dx = \int_{\partial U} \left(u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n}\right) dS \tag{3.2}$$

for $u, v \in C^2(\overline{U})$.

## 3.2 Mollification and smoothing

To approximate an arbitrary function by a smooth one we often resort to convolution, as follows.

Consider a **bump function** $\eta(x)$ on $\mathbb{R}^N$, by which we mean $\eta$ is smooth, radial, nonnegative, and supported in the unit ball $B$, with $\int_B \eta\,dx = 1$. For example, one could take

$$\eta(x) = C \begin{cases} \exp\left(1/(|x|^2 - 1)\right), & |x| < 1, \\ 0, & |x| \geq 1, \end{cases}$$

where the constant $C$ is used to normalize the integral of $\eta$. Let

$$\eta_\varepsilon(x) = \varepsilon^{-N}\eta(\varepsilon^{-1}x),$$

so that $\eta_\varepsilon$ is supported in the ball of radius $\varepsilon$, and has integral $1$ over that ball. Intuitively, $\eta_\varepsilon$ converges to a delta function as $\varepsilon \to 0$. The theorem below makes that intuition precise.

The **mollification** of a locally integrable function $f$ on $U$ is the convolution

$$f^\varepsilon = \eta_\varepsilon * f.$$

Thus $f^\varepsilon(x)$ equals an average of values of $f$ near the point $x$:

$$\begin{aligned} f^\varepsilon(x) &= \int_{B(\varepsilon)} \eta_\varepsilon(y) f(x-y)\,dy \\ &= \int_U \eta_\varepsilon(x-y) f(y)\,dy \end{aligned}$$

where $f^\varepsilon(x)$ is well defined for $x$ belonging to the set

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

**Theorem 3.1** (Mollification)**.** *Let* $f \in L^1_{loc}(U)$.
*(a) The function* $f^\varepsilon = \eta_\varepsilon * f$ *is smooth on* $U_\varepsilon$, *with* $D^\alpha f^\varepsilon = (D^\alpha \eta_\varepsilon) * f$.
*(b) [Pointwise convergence]* $f^\varepsilon(x) \to f(x)$ *as* $\varepsilon \to 0$, *for almost every* $x \in U$, *or more precisely, for every Lebesgue point of* $f$.
*(c) [*$L^\infty$*-convergence] If* $f$ *is continuous, then on each compact subset of* $U$ *we have* $f^\varepsilon \to f$ *uniformly, as* $\varepsilon \to 0$.
*(d) [*$L^p$*-convergence] If* $1 \le p < \infty$ *and* $f \in L^p_{loc}(U)$ *then* $f^\varepsilon \to f$ *in* $L^p_{loc}(U)$, *as* $\varepsilon \to 0$.

The notation $D^\alpha$ for the $\alpha$-th derivative is defined in the next section.

*Proof.* Part (a) Fix $\varepsilon > 0$ and $x \in U_\varepsilon$, so that $\varepsilon < \operatorname{dist}(x, \partial U)$. Choose $\delta > 0$ small enough that $\varepsilon + 2\delta < \operatorname{dist}(x, \partial U)$. Fix $j \in \{1, \dots, N\}$ and consider real numbers $h \in (-\delta, \delta)$. Then we can evaluate the difference quotient for $f^\varepsilon$ as

$$\frac{f^\varepsilon(x + he_j) - f^\varepsilon(x)}{h} = \frac{1}{\varepsilon^N} \int_{U_\delta} \frac{1}{h}\Big[\eta\big(\frac{x + he_j - y}{\varepsilon}\big) - \eta\Big(\frac{x-y}{\varepsilon}\Big)\Big] f(y)\,dy,$$

where on the right side we need only integrate over $y \in U_\delta$ because if $y \notin U_\delta$ then $|x - y| > \varepsilon$ and $|x + he_j - y| > \varepsilon$ and so the terms with $\eta$ would vanish. Next, we know $f$ is integrable on $U_\delta$ and

$$\frac{1}{h}\Big[\eta\big(\frac{x + he_j - y}{\varepsilon}\big) - \eta\Big(\frac{x-y}{\varepsilon}\Big)\Big] \to \frac{1}{\varepsilon}\frac{\partial \eta}{\partial x_j}\Big(\frac{x-y}{\varepsilon}\Big)$$

as $h \to 0$, uniformly for $y \in \mathbb{R}^N$ (because $\eta$ is smooth with compact support). Hence by letting $h \to 0$ in the difference quotient we conclude (using Exercise 3.1 below) that the partial derivative $\partial f^\varepsilon / \partial x_j$ exists at $x$ and equals

$$\int_U \frac{\partial \eta_\varepsilon}{\partial x_j}(x - y) f(y)\,dy.$$

By repeating the argument we obtain the formula for $D^\alpha f^\varepsilon$.

Part (b). Consider a Lebesgue point $x$ for $f$. Then

$$\begin{aligned}
|f^\varepsilon(x) - f(x)| &= \Big| \int_U \eta_\varepsilon(x - y)[f(y) - f(x)]\,dy \Big| \qquad \text{since } \eta_\varepsilon \text{ has integral } 1 \\
&\le \frac{1}{\varepsilon^N} \int_{B(x,\varepsilon)} \eta\Big(\frac{x-y}{\varepsilon}\Big)|f(y) - f(x)|\,dy \\
&\le |B(1)|\|\eta\|_{L^\infty} \frac{1}{|B(x,\varepsilon)|} \int_{B(x,\varepsilon)} |f(y) - f(x)|\,dy \\
&\to 0 \qquad (3.3)
\end{aligned}$$

as $\varepsilon \to 0$, by the Lebesgue Differentiation Theorem.

Part (c). Suppose $f$ is continuous. Given an arbitrary domain $W \Subset U$ we choose $V$ such that $W \Subset V \Subset U$. Note $f$ is uniformly continuous on the compact set $\overline{V}$. Hence the limit (3.3) holds uniformly for $x \in W$, and so $f^\varepsilon \to f$ uniformly on $W$.

Part (d). Let $f \in L^p_{loc}(U)$, and consider arbitrary subdomains $W \Subset V \Subset U$, so that $f \in L^p(V)$. We start by showing that the norm of the mollified function on $W$ is bounded by the norm of the original function on $V$. Extend $f$ from $V$ to all of $\mathbb{R}^N$ by letting

$$F = \begin{cases} f & \text{on } V, \\ 0 & \text{off } V. \end{cases}$$

Then on $W$ we have $f^\varepsilon = \eta_\varepsilon * F$ whenever $\varepsilon < \text{dist}(W, \partial V)$, and so

$$\begin{aligned} \|f^\varepsilon\|_{L^p(W)} &\leq \|\eta_\varepsilon * F\|_{L^p(\mathbb{R}^N)} \\ &\leq \|\eta_\varepsilon\|_{L^1(\mathbb{R}^N)} \|F\|_{L^p(\mathbb{R}^N)} \\ &= \|f\|_{L^p(V)}. \end{aligned} \tag{3.4}$$

To complete the proof, let $\beta > 0$ and choose a continuous function $g$ on $V$ such that $\|f - g\|_{L^p(V)} < \beta$. (Here we need that $p \neq \infty$.) Then

$$\begin{aligned} \|f^\varepsilon - f\|_{L^p(W)} &\leq \|(f-g)^\varepsilon\|_{L^p(W)} + \|g^\varepsilon - g\|_{L^p(W)} + \|g - f\|_{L^p(W)} \\ &\leq \|f - g\|_{L^p(V)} + \|g^\varepsilon - g\|_{L^p(W)} + \beta \qquad \text{by (3.4)} \\ &\leq 2\beta + o(1) \end{aligned}$$

since $g^\varepsilon \to g$ uniformly on $W$ as $\varepsilon \to 0$. Hence $\limsup_{\varepsilon\to 0} \|f^\varepsilon - f\|_{L^p(W)} \leq 2\beta$, and letting $\beta \to 0$ shows that $\lim_{\varepsilon\to 0} \|f^\varepsilon - f\|_{L^p(W)} = 0$, as desired. □

***Exercise*** 3.1. Let $(X, \mu)$ be a measure space. Show that if $f_n \to f$ in $L^\infty(X, \mu)$ and $g \in L^1(X, \mu)$ then $\int_X f_n g \, d\mu \to \int_X fg \, d\mu$ as $n \to \infty$.

***Exercise*** 3.2. Deduce from Theorem 3.1 that if $\nu$ is locally integrable and $\int_U \nu(y)\phi(y)\, dy = 0$ for all $\phi \in C_c^\infty(U)$, then $\nu = 0$ a.e. in $U$.

## 3.3 Weak derivatives and Sobolev spaces

A **multiindex** is a vector $\alpha = (\alpha_1, \ldots, \alpha_N)$ with nonnegative integer entries. Write

$$D^\alpha \phi = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_N}}{\partial x_N^{\alpha_N}} \phi$$

for the $\alpha$-th derivative of the smooth function $\phi$. By convention, $D^0\phi = \phi$. Notice $D^\alpha\phi$ is a derivative of **order**

$$|\alpha| = \alpha_1 + \cdots + \alpha_N.$$

(One uses the $\ell^1$-norm on multiindices, rather than the $\ell^2$-norm, because the $\ell^1$-norm helpfully computes the total order of the derivative $D^\alpha\phi$.)

**Definition 3.2** (Weak derivatives)**.** Let $u$ and $v$ be locally integrable functions on $U$. We say $D^\alpha u = v$ **weakly** if

$$\int_U u D^\alpha \phi \, dx = (-1)^{|\alpha|} \int_U v\phi \, dx \qquad \forall \phi \in C_c^\infty(U). \tag{3.5}$$

Notice $u$ and $v$ can be changed on a set of measure zero without affecting this condition.

For example, in $1$ dimension, $u' = v$ weakly on $U = (a, b)$ if

$$\int_a^b u\phi' \, dx = -\int_a^b v\phi \, dx. \tag{3.6}$$

To motivate the definition, note that if $u' = v$ classically then equation (3.6) is simply what you get after integration by parts on the left side, using the compact support of $\phi$ to insure that $\phi(a) = 0$ and $\phi(b) = 0$. Similarly, if $D^\alpha u = v$ classically then (3.5) is what you get after integrating by parts $\alpha$ times on the left side.

The point of this definition of weak derivative is that (3.5) and (3.6) can hold even when $u$ is *not* classically differentiable at some points, as the next example shows.

***Exercise*** 3.3. In one dimension, let

$$u(x) = \begin{cases} x, & 0 < x \leq 1, \\ 2 - x, & 1 \leq x < 2, \\ 0, & \text{otherwise}, \end{cases} \qquad v(x) = \begin{cases} 1, & 0 < x < 1, \\ -1, & 1 < x < 2, \\ 0, & \text{otherwise}. \end{cases}$$

Show that $u' = v$ weakly on $\mathbb{R}$, even though the graph of $u$ has a corner at $x = 1$. *Note.* Your proof will need to use continuity of $u$ at $x = 1$.

***Exercise*** 3.4. Construct a weakly differentiable function $u$ on $\mathbb{R}$ whose graph has infinitely many corners on the unit interval $[0, 1]$.

**Lemma 3.3** (Classical derivatives are also weak derivatives). *Suppose* $u$ *is smooth, and let* $v = D^{\alpha}u$ *denote the classical* $\alpha$*-th derivative of* $u$*. Then* (3.5) *holds, so that* $D^{\alpha}u = v$ *weakly.*

***Exercise*** 3.5. Prove Lemma 3.3, using integration by parts.

**Lemma 3.4** (Weak derivatives are unique). *Assume the weak derivative* $D^{\alpha}u$ *exists. Then it is unique, up to a set of measure zero.*

***Exercise*** 3.6. Prove Lemma 3.4.

Now we can define the Sobolev space $W^{k,p}$ as the collection of functions in $L^p$ whose derivatives up to order $k$ also belong to $L^p$. Here the "$W$" refers to the "weak" nature of the derivatives.

**Definition 3.5** (Sobolev spaces). For $1 \le p \le \infty$ and $k \ge 0$, let

$$W^{k,p}(U) = \big\{u \in L^1_{loc}(U) : D^{\alpha}u \text{ exists weakly whenever } 0 \le |\alpha| \le k, \\ \text{with } D^{\alpha}u \in L^p(U)\big\}.$$

This definition makes sense on any open set $U$, regardless of whether $U$ is connected or bounded.

In the special case $p = 2$ we write

$$\boxed{H^k = W^{k,2}}.$$

**Theorem 3.6.**
*(i)* $W^{k,p}(U)$ *is a Banach space using the norm*

$$\|u\|_{W^{k,p}} = \Big(\sum_{|\alpha| \le k} \|D^{\alpha}u\|_{L^p}^p\Big)^{1/p} = \Big(\int_U \sum_{|\alpha| \le k} |D^{\alpha}u|^p \, dx\Big)^{1/p}$$

*when* $1 \le p < \infty$, *and* $\|u\|_{W^{k,\infty}} = \sum_{|\alpha| \le k} \|D^{\alpha}u\|_{L^{\infty}}$ *when* $p = \infty$.
*(ii)* $H^k(U)$ *is a Hilbert space with inner product*

$$\langle u, v\rangle_{H^k} = \sum_{|\alpha| \le k} \langle D^{\alpha}u, D^{\alpha}v\rangle_{L^2} = \int_U \sum_{|\alpha| \le k} D^{\alpha}u \, D^{\alpha}v \, dx.$$

*In particular, taking* $k = 1$ *and* $p = 2$ *gives the* $W^{1,2}$*-inner product:*

$$\langle u, v\rangle_{H^1} = \langle u, v\rangle_{W^{1,2}} = \int_U (uv + \nabla u \cdot \nabla v) \, dx.$$

*Proof.* First one checks that weak derivatives are linear, meaning $D^\alpha(u_1 + u_2) = D^\alpha u_1 + D^\alpha u_2$ and so on. Hence $W^{k,p}$ is a linear space. That the definition above indeed gives a norm is easily checked, except that the triangle inequality requires careful proof. That $\langle \cdot, \cdot \rangle_{H^k}$ is an inner product is essentially immediate.

To prove completeness, suppose $\{u_n\}$ is Cauchy in $W^{k,p}(U)$. Then

$$\|u_n - u_m\|_{W^{k,p}}^p = \sum_{|\alpha| \leq k} \|D^\alpha u_n - D^\alpha u_m\|_{L^p}^p \to 0$$

as $n, m \to \infty$. Hence $\{D^\alpha u_n\}$ is Cauchy in $L^p(U)$, for each $\alpha$, and thus it converges to a limit, say

$$u_\alpha = \lim_{n \to \infty} D^\alpha u_n \qquad \text{in } L^p(U). \tag{3.7}$$

When $\alpha = 0$, write $u = \lim_n u_n$.

We will show that $D^\alpha u = u_\alpha$ weakly whenever $0 < |\alpha| \leq k$, so that $u \in W^{k,p}(U)$ and (3.7) implies $u_n \to u$ in the $W^{k,p}$-norm, thus proving completeness.

To show $D^\alpha u = u_\alpha$ weakly, we observe that for all $\phi \in C_c^\infty(U)$,

$$\begin{aligned}
\int_U u D^\alpha \phi \, dx &= \lim_n \int_U u_n D^\alpha \phi \, dx && \text{by Hölder, since } u_n \to u \text{ in } L^p(U), \\
&= \lim_n (-1)^{|\alpha|} \int_U (D^\alpha u_n) \phi \, dx && \text{since } D^\alpha u_n \text{ exists weakly,} \\
&= (-1)^{|\alpha|} \int_U u_\alpha \phi \, dx && \text{by Hölder, since } D^\alpha u_n \to u_\alpha \text{ in } L^p(U).
\end{aligned}$$

□

***Exercise*** 3.7 (Basic properties).
(i) Weak derivatives commute: $D^\alpha D^\beta u = D^\beta D^\alpha u$, assuming all these weak derivatives exist individually.
(ii) If $\zeta \in C_c^\infty(U)$ and $u \in W^{k,p}(U)$ then $\zeta u \in W^{k,p}(U)$ and its derivatives are given by the product rule (Leibniz formula).

***Exercise*** 3.8 (Singular Sobolev function in dimension bigger than $1$). Let $N \geq 2$ and $1 \leq p < N$. Fix $q$ with

$$0 < q < \frac{N - p}{p}.$$

Then $|x|^{-q} \in W^{1,p}(B)$ where $B$ is the unit ball.

Thus a Sobolev function can have an unbounded singularity. Even worse, it can have singularities all over the place:

***Exercise*** 3.9. Construct a Sobolev function whose singularities form a dense subset of the unit ball $B$, when $N \geq 2$. For simplicity, you may take $N = 2, p = 1, 0 < q < 1$.

Sobolev functions in $1$ dimension are better behaved: singularities cannot occur, as shown by the exercises in the next section.

## 3.4 Approximating Sobolev functions by smooth functions

We will prove results for Sobolev functions by first treating smooth functions and then passing to a limit, with the help of density results in this section.

**Proposition 3.7** (Local approximation). *Assume* $u \in W^{k,p}(U), 1 \leq p < \infty$. *Then* $u^\varepsilon \to u$ *in* $W^{k,p}_{loc}(U)$ *as* $\varepsilon \to 0$.

*Proof.* The mollification $u^\varepsilon = \eta_\varepsilon * u$ is smooth by Theorem 3.1(a). We now show it has classical derivative

$$D^\alpha(\eta_\varepsilon * u) = \eta_\varepsilon * D^\alpha u$$

where $D^\alpha u$ is the weak derivative of $u$, for $|\alpha| \leq k$. Indeed, for $x \in U_\varepsilon$,

$$\begin{aligned}
(\eta_\varepsilon * D^\alpha u)(x) &= \int_U \eta_\varepsilon(x-y) D^\alpha u(y)\, dy \\
&= \int_U (-1)^{|\alpha|} D^\alpha_y \big(\eta_\varepsilon(x-y)\big) u(y)\, dy \qquad \text{by definition of weak} \\
&\qquad\qquad \text{derivative, with test function } \phi(y) = \eta_\varepsilon(x-y), \\
&= \int_U (D^\alpha \eta_\varepsilon)(x-y) u(y)\, dy \qquad \text{by the chain rule} \\
&= (D^\alpha \eta_\varepsilon) * u(x) \\
&= D^\alpha(\eta_\varepsilon * u)(x) \qquad \text{by Theorem 3.1(a) (diff. through the integral)} \\
&= D^\alpha u^\varepsilon(x).
\end{aligned}$$

Hence $D^\alpha u^\varepsilon = \eta_\varepsilon * D^\alpha u \to D^\alpha u$ in $L^p_{loc}(U)$ by Theorem 3.1(d), as $\varepsilon \to 0$. This holds for each index $\alpha$, and so $u^\varepsilon \to u$ in $W^{k,p}_{loc}(U)$. □

***Exercise*** 3.10 (Fundamental theorem for Sobolev functions). Suppose $u \in W^{1,p}(a,b)$ for some $p \geq 1$, and that $c$ is a Lebesgue point of $u$.
(i) Show that

$$u(x) = u(c) + \int_c^x u'(z)\,dz$$

for almost every $x \in (a,b)$. Here $u'$ is the weak derivative of $u$.
(ii) Deduce that if $u(c) = 0$ for some Lebesgue point $c$ then

$$\|u\|_{L^\infty(a,b)} \leq \|u'\|_{L^1(a,b)}.$$

(iii) Define $\widetilde{u}(x)$ to be the right side in part (i), so that $\widetilde{u} = u$ a.e. Show $\widetilde{u}$ is absolutely continuous. (Thus Sobolev functions in $1$ dimension are identified with absolutely continuous functions having a.e.-classical derivative in $L^p$.)

***Exercise*** 3.11 (Sobolev implies Hölder continuity, in $1$ dimension). Suppose $u \in W^{1,p}(a,b)$ where $1 < p \leq \infty$. Show that (after redefining $u$ on a set of measure zero) we have

$$|u(x) - u(y)| \leq \|u'\|_{L^p(a,b)} |x-y|^{1-1/p}, \qquad x, y \in (a,b).$$

Hence $u$ is Hölder continuous with exponent $1 - 1/p$. (When $p = \infty$, this means $u$ is Lipschitz continuous.)

***Exercise*** 3.12 (Vanishing weak derivative implies the function is constant). Show that if $u$ has weak derivative zero in each coordinate direction ($\nabla u \equiv 0$ weakly) then $u$ is constant a.e. *Note.* Your proof must at some stage use connectedness of the domain $U \subset \mathbb{R}^N$.

Next we prove smooth functions are dense in $W^{k,p}$.

**Theorem 3.8** (Global approximation by smooth functions)**.** *If* $u \in W^{k,p}(U)$, $1 \leq p < \infty$, *then functions* $u_n \in C^\infty \cap W^{k,p}(U)$ *exist such that* $u_n \to u$ *in* $W^{k,p}(U)$.

*Proof.* Use the local approximation result in Proposition 3.7 together with a partition of unity exhaustion of the domain. See [Evans, Chapter 5]. □

**Theorem 3.9** (Global approximation by functions smooth up to the boundary)**.** *Assume* $\partial U$ *is* $C^1$*-smooth (see definition below). If* $u \in W^{k,p}(U)$, $1 \leq p < \infty$, *then functions* $u_n \in C^\infty(\overline{U})$ *exist such that* $u_n \to u$ *in* $W^{k,p}(U)$.

*Proof.* Use the global approximation from Theorem 3.8 to reduce to the case of $u \in C^\infty(U)$. Then "translate $u$ outwards across the boundary" to get an approximating function in $C^\infty(\overline{U})$. See [Evans, Chapter 5]. □

**Definition 3.10** (Smoothness of the boundary)**.** The boundary $\partial U$ is called $C^m$**-smooth** if for each $x^* \in \partial U$ there exists a choice of coordinate system, a radius $r > 0$, and a $C^m$-smooth function $\gamma : \mathbb{R}^{N-1} \to \mathbb{R}$, such that

$$U \cap B(x^*, r) = \{x \in B(x^*, r) : x_N > \gamma(x_1, \ldots, x_{N-1})\}. \tag{3.8}$$

Thus the boundary is expressed as the graph of a $C^m$-smooth function, and the domain lies above the graph.

*Note.* In $1$ dimension, an interval $U$ is always regarded as having smooth boundary.

## 3.5 Sobolev space of functions vanishing on the boundary

**Definition 3.11.** Define $W_0^{k,p}(U)$ to be the closure in $W^{k,p}(U)$ of the subspace $C_c^\infty(U)$ of smooth functions having compact support in $U$. Thus $W_0^{k,p}$ is a Banach space, since it is a closed subspace of a Banach space. When $p = 2$ we write

$$H_0^k = W_0^{k,2}.$$

This abstract definition in terms of closures can be difficult to understand, the first time it is encountered. In explicit terms, the definition says that $u \in W_0^{k,p}(U)$ if $u \in W^{k,p}(U)$ and there exists a sequence of compactly supported functions $u_n \in C_c^\infty(U)$ such that $u_n \to u$ in $W^{k,p}(U)$. Remember that convergence in the Sobolev norm means:

$$u_n \to u \text{ and } D^\alpha u_n \to D^\alpha u \text{ in } L^p(U), \text{ for each multiindex with } |\alpha| \le k.$$

Thus global approximation with smooth functions holds by definition for $W_0^{k,p}(U)$ — unlike for $W^{k,p}(U)$, where global approximation had to be proved, in Theorems 3.8 and 3.9.

For the special case $k = 1$ and $p = 2$, the definition means that $u \in H_0^1(U)$ if $u \in H^1(U)$ and there exists a sequence of compactly supported functions $u_n \in C_c^\infty(U)$ such that $u_n \to u$ in $H^1(U)$. This norm convergence means

$$u_n \to u \text{ and } \nabla u_n \to \nabla u \text{ in } L^2(U).$$

***Exercise*** 3.13. Prove that

$$\int_U u^2\,dx \le C\int_U |\nabla u|^2\,dx, \qquad u \in H^1_0(U),$$

for some constant $C$ depending on the domain $U$.

***Exercise*** 3.14 (Weak solution of Poisson's equation). In Section 2.2 we outlined an argument for finding a weak solution of Poisson's equation. Make that argument rigorous: prove that if $f \in L^2(U)$ then a unique function $u \in H^1_0(U)$ exists such that $-\Delta u = f$ weakly in $H^1_0(U)$.

## 3.6 Extending past the boundary

Sobolev functions can be extended to the *exterior* of the domain while increasing the Sobolev norm by at most a bounded factor.

**Theorem 3.12** (Extension operator for $p < \infty$)**.** *Assume* $1 \le p < \infty$*. Let* $U$ *and* $V$ *be bounded domains in* $\mathbb{R}^N$ *with* $U \Subset V$ *and suppose* $\partial U$ *is* $C^1$*-smooth.*

*Then there exists a bounded linear operator*

$$E : W^{1,p}(U) \to W^{1,p}_0(V) \subset W^{1,p}(\mathbb{R}^N)$$

*such that*

- $Eu = u$ *a.e. in* $U$*,*
- supp$(Eu) \subset V$*,*
- $\|Eu\|_{W^{1,p}(\mathbb{R}^N)} \le C\|u\|_{W^{1,p}(U)}$ *for some constant* $C = C(N, U, V)$*.*

*Proof.* The idea is to straighten the boundary, then reflect the function across it, and finally apply a cut-off function. Assume $u \in C^1(\overline{U})$, until the final step in the proof.

Step 1 — Reflection across a flat boundary. Fix $x^* \in \partial U$ and assume there exists a choice of coordinate system and a radius $r^* > 0$ such that $x^* = 0$ and

$$U \cap B = \{x \in B : x_N > 0\} \equiv B^+,$$
$$U^c \cap B = \{x \in B : x_N \le 0\} \equiv B^-,$$

where $B = B(x^*, r^*)$. We know $u \in C^1(\overline{B^+})$. Extend $u$ across the flat boundary by defining

$$\overline{u}(x) = \begin{cases} u(x) & \text{in } \overline{B^+} \\ -3u(x', -x_N) + 4u(x', -x_N/2) & \text{in } \overline{B^-} \end{cases}$$

where we use the notation $x' = (x_1, \ldots, x_{N-1})$. Then

- $\overline{u}$ is continuous at the boundary ($x_N = 0$), since $-3 + 4 = 1$;
- similarly $\overline{u}_{x_j}$ is continuous at the boundary, for $j = 1, \ldots, N-1$;
- $\overline{u}_{x_N}$ is continuous when $x_N = 0$, since $(-3)(-1) + 4(-1/2) = 1$.

Hence $\overline{u} \in C^1(\overline{B})$. You can check that

$$\|\overline{u}\|_{W^{1,p}(B)} \le C\|u\|_{W^{1,p}(B^+)}$$

where $C$ is an absolute constant.

Step 2 — Straightening the boundary (in dimensions $N \ge 2$). Drop the assumption that the boundary is flat near $x^*$. Instead, take the function $\gamma$ in the definition (3.8) for $C^1$-boundary, and define a $C^1$-diffeomorphism $\Phi : \mathbb{R}^N \to \mathbb{R}^N$ by $\Phi(x) = y$ where

$$\begin{aligned} y_N &= x_N - \gamma(x'), \\ y_j &= x_j, \qquad \text{for } j = 1, \ldots, N-1. \end{aligned}$$

Let $y^* = \Phi(x^*)$ and choose $r^* > 0$ small enough that $B(y^*, r^*) \subset \Phi(B(x^*, r))$. Let $B = B(y^*, r^*)$. Then from (3.8) we deduce that

$$\Phi(U) \cap B = \{y \in B : y_N > 0\}.$$

Let $v = u \circ \Phi^{-1} \in C^1(\overline{B^+})$ and extend by reflection to $\overline{v} \in C^1(\overline{B})$ as in Step 1, so that

$$\|\overline{v}\|_{W^{1,p}(B)} \le C\|v\|_{W^{1,p}(B^+)}$$

To undo the straightening of the boundary we define $\overline{u} = \overline{v} \circ \Phi$ on the neighborhood $W = \Phi^{-1}(B)$ of $x^*$. By substituting $y = \Phi(x)$ in the preceding estimate and using the chain rule suitably, we deduce

$$\|\overline{u}\|_{W^{1,p}(W)} \le C\|u\|_{W^{1,p}(W^+)} \tag{3.9}$$

for some new constant $C = C(x^*, \Phi)$, where $W^+ = \Phi^{-1}(B^+)$.

Note we can suppose $W \Subset V$, by taking $r^*$ smaller if necessary.

Step 3 — Localization. The boundary of $U$ is compact, and so we may cover it with finitely many neighborhoods $W_i$ of the kind found in Step 2.

Choose an additional open set $W_0 \Subset U$ that covers all points in $U$ that are not covered by the other $W_i$. Let $\overline{u} = u$ on $W_0$.

One may "piece together" the different extensions using a partition of unity subordinate to the $W_i$, and hence obtain a global extension of $u$, which we call $Eu$; see [Evans, Section 5.4] for details. Then from (3.9) we obtain the norm estimate on $Eu$ that is stated in the theorem.

Step 4 — Using density. Clearly $E : C^1(\overline{U}) \to W^{1,p}(\mathbb{R}^N)$ is linear, because the reflection construction in Step 1 is linear. And the operator $E$ is bounded with respect to the Sobolev norms, by (3.9). Since $C^1(\overline{U})$ is dense in $W^{1,p}(U)$ by Theorem 3.9, the operator $E$ extends to a bounded linear operator from $W^{1,p}(U)$ to $W^{1,p}(\mathbb{R}^N)$ (see Exercise 1.18).

Finally, $Eu$ is compactly supported in $V$ because the partition of unity is supported in $V$. Hence by mollification, $Eu$ can be approximated in the Sobolev norm by smooth functions with compact support in $V$. Thus $Eu \in W_0^{1,p}(V)$. □

**Corollary 3.13** (Extension for $p = \infty$)**.** *The Extension Theorem 3.12 remains valid for* $p = \infty$*, except we do not claim that* $E$ *maps into* $W_0^{1,\infty}(V)$*; rather we have* $E : W^{1,\infty}(U) \to W^{1,\infty}(\mathbb{R}^N)$*.*

*Proof.* One simply lets $p \to \infty$ in the preceding theorem. Here we use that if $u \in W^{1,\infty}(U)$ then $u \in W^{1,p}(U)$, and that for any function $f$ we have $\|f\|_{L^p(U)} \to \|f\|_{L^\infty(U)}$ as $p \to \infty$, since $U$ has finite measure. □

## 3.7 Boundary traces

Functions in $L^p$ need not have boundary values. First of all, the boundary has measure zero and hence we may change the values there without affecting the function as an element of $L^p$. Second, even for smooth functions the boundary values might be identically infinite; a simple example is

$$u(x) = (1 - |x|)^{-1/4}, \qquad |x| < 1,$$

which is square integrable on the unit ball.

Fortunately, Sobolev functions do have boundary values thanks to the $L^p$ control on the derivative, as the next theorem shows.

**Theorem 3.14** (Boundary trace operator). *Assume* $\partial U$ *is* $C^1$*-smooth and* $1 \le p < \infty$*. Then a linear operator*

$$T : W^{1,p}(U) \to L^p(\partial U)$$

*exists such that* $Tu = u|_{\partial U}$ *whenever* $u \in C(\overline{U}) \cap W^{1,p}(U)$*. (Thus* $T$ *captures the boundary values of continuous functions.) The operator* $T$ *is bounded, with*

$$\|Tu\|_{L^p(\partial U)} \le C\|u\|_{W^{1,p}(U)}, \qquad u \in W^{1,p}(U), \tag{3.10}$$

*where* $C = C(p, U)$*.*

*Proof.* Suppose $u \in C^1(\overline{U})$. Since this function has well-defined boundary values, we can let

$$Tu = u|_{\partial U}.$$

Clearly $T : C^1(\overline{U}) \to L^p(\partial U)$ is a linear operator.

Step 1 — Flat boundary. Let $x^* \in \partial U, r > 0, B = B(x^*, r)$, and assume $U \cap B = B^+$ and $U^c \cap B = B^-$, just as in the previous proof. Write

$$\Gamma = \left\{x \in B(x^*, r/2) : x_N = 0\right\} = B(x^*, r/2) \cap \partial U.$$

Choose $\zeta(x)$ to be a smooth, nonnegative cut-off function with $\zeta \equiv 1$ on $B(x^*, r/2)$ and $\zeta \equiv 0$ outside $B(x^*, r)$.

To control the boundary values of $u$ on $\Gamma$ we observe

$$\begin{aligned}
\int_\Gamma |Tu|^p \, dx' &\le \int_{\{x_N=0\}} \zeta |u|^p \, dx' && \text{since } \zeta \equiv 1 \text{ on } \Gamma \\
&= -\int_{B^+} \left(\zeta |u|^p\right)_{x_N} dx && \text{by fundamental theorem in } x_N\text{-variable} \\
&= -\int_{B^+} \left(\zeta_{x_N} |u|^p + p|u|^{p-1} \operatorname{sign}(u) u_{x_N}\right) dx \\
&\le C \int_{B^+} \left(|u|^p + |u|^{p-1} |\nabla u|\right) dx \\
&\le C \int_{B^+} \left(|u|^p + |\nabla u|^p\right) dx
\end{aligned}$$

by the elementary inequality $a^{p-1}b \le a^p + b^p$; here the constant $C$ depends on $p$ and on $x^*$ (since $C$ depends on $\zeta$).

Step 2 — Curved boundary (in dimensions $N \ge 2$). If the boundary is not flat near $x^*$ then by straightening the boundary with a $C^1$-diffeomorphism we deduce from Step 1 that

$$\int_\Gamma |u|^p \, dS \le C \int_U (|u|^p + |\nabla u|^p) \, dx$$

where now $\Gamma$ is an open subset of $\partial U$ containing the point $x^* \in \partial U$. Note that $C = C(p, x^*)$. By covering the boundary with finitely many such balls, we deduce

$$\int_{\partial U} |u|^p \, dS \le C \int_U (|u|^p + |\nabla u|^p) \, dx$$

where now $C$ depends on $p$ and $U$. Hence (3.10) holds for $u \in C^1(\overline{U})$.

Step 3 — Extending to Sobolev space. The operator $T$ extends to a bounded linear operator on $u \in W^{1,p}(U)$, by (3.10) and density of $C^1(\overline{U})$.

Step 4 — Boundary values of continuous Sobolev functions. If $u \in C(\overline{U}) \cap W^{1,p}(U)$, with $u$ not necessarily $C^1$-smooth, then $Tu$ still equals $u$ on the boundary; see [Evans, Chapter 5]. □

For partial differential equations, we often want to impose zero Dirichlet boundary conditions. The next corollary shows the boundary condition is equivalent to the solution belonging to $W_0^{1,p}$.

**Corollary 3.15** (Zero trace iff $W_0^{1,p}$). *Assume $\partial U$ is $C^1$-smooth and $u \in W^{1,p}(U)$ with $1 \le p < \infty$. Then*

$$Tu = 0 \text{ on } \partial U \qquad \Longleftrightarrow \qquad u \in W_0^{1,p}(U).$$

*Proof.* "$\Longleftarrow$" If $u \in W_0^{1,p}(U)$ then $u = \lim u_n$ in the $W^{1,p}$-norm, for some smooth functions $u_n$ with compact support in $U$. Hence

$$Tu = \lim Tu_n = 0$$

in $L^p(\partial U)$, where we use that $Tu_n = 0$ (because $u_n \in C^1(\overline{U})$ and $u_n = 0$ on the boundary) and that $T$ is a continuous operator.

"$\Longrightarrow$" See [Evans, Chapter 5]. □

## 3.8 Sobolev inequalities

If $u \in W^{1,p}$ then of course $u \in L^p$. We will prove the better result that $u \in L^{p^*}$ for some $p^* > p$. That is, control over the derivatives of $u$ gives higher integrability of $u$ itself.

For example, in 1 dimension $W^{1,1} \subset L^\infty$ by the fundamental theorem of calculus (Exercise 3.10).

**Scaling argument.**

We start by guessing the Gagliardo–Nirenberg–Sobolev (GNS) inequality for functions on $\mathbb{R}^N$. Fix $p, q > 0$ and assume that

$$\|u\|_{L^q(\mathbb{R}^N)} \leq C\|\nabla u\|_{L^p(\mathbb{R}^N)}, \qquad u \in C^1_c(\mathbb{R}^N), \tag{3.11}$$

for some $C > 0$. Fix the function $u \not\equiv 0$, and rescale it to obtain the function $u_\lambda(x) = u(\lambda x)$ where $\lambda > 0$ is constant. Applying (3.11) to $u_\lambda$ gives

$$\Big(\int |u(\lambda x)|^q \, dx\Big)^{1/q} \leq C\Big(\int |\lambda(\nabla u)(\lambda x)|^p \, dx\Big)^{1/p}.$$

We change variable with $y = \lambda x$ and deduce that

$$0 < \|u\|_{L^q(\mathbb{R}^N)} \leq \lambda^{1+N(q^{-1}-p^{-1})} C\|\nabla u\|_{L^p(\mathbb{R}^N)}.$$

If the exponent of $\lambda$ were positive then we would obtain a contradiction by letting $\lambda \to 0$, and if it were negative we could similarly let $\lambda \to \infty$. Hence the exponent must equal zero, which says

$$\frac{1}{q} = \frac{1}{p} - \frac{1}{N}.$$

We have shown this condition is necessary for (3.11) to hold. The theorem below shows it is sufficient.

**Definition 3.16.** For $N \geq 2$, the **Sobolev conjugate** of $p \in [1, N)$ is

$$p^* = \frac{Np}{N-p}, \qquad \text{so that} \qquad \frac{1}{p^*} = \frac{1}{p} - \frac{1}{N}.$$

For $N = 1$, we let $1^* = \infty$.

Notice $p < p^*$, and that if $p \nearrow N$ then $p^* \nearrow \infty$.

**Theorem 3.17** (Gagliardo–Nirenberg–Sobolev inequality)**.** *If* $N = 1, p = 1$, *then*

$$\|u\|_{L^\infty(\mathbb{R})} \le \|u'\|_{L^1(\mathbb{R})}, \qquad u \in C_c^1(\mathbb{R}).$$

*If* $N \ge 2$ *and* $p \in [1, N)$ *then*

$$\|u\|_{L^{p^*}(\mathbb{R}^N)} \le C\|\nabla u\|_{L^p(\mathbb{R}^N)}, \qquad u \in C_c^1(\mathbb{R}^N),$$

*where* $C = C(p, N)$.

*Proof.* Assume $N = 1, p = 1$, and $u \in C_c^1(\mathbb{R})$ . By the fundamental theorem,

$$|u(x)| = \Big|\int_{-\infty}^{x} u'(y)\, dy\Big| \le \|u'\|_{L^1(\mathbb{R})},$$

giving the result.

Now assume $N = 2$ and $p = 1$, so that $p^* = 2$. Let $u \in C_c^1(\mathbb{R}^2)$. By the fundamental theorem,

$$|u(x_1, x_2)| = \Big|\int_{-\infty}^{x_1} u_{y_1}(y_1, x_2)\, dy_1\Big| \le \int_{\mathbb{R}} |\nabla u(y_1, x_2)|\, dy_1,$$
$$|u(x_1, x_2)| = \Big|\int_{-\infty}^{x_2} u_{y_2}(x_1, y_2)\, dy_2\Big| \le \int_{\mathbb{R}} |\nabla u(x_1, y_2)|\, dy_2.$$

Multiplying and integrating gives that

$$\int_{\mathbb{R}^2} |u(x_1, x_2)|^2\, dx_1 dx_2 \le \int_{\mathbb{R}^2} |\nabla u(y_1, x_2)|\, dy_1 dx_2 \int_{\mathbb{R}^2} |\nabla u(x_1, y_2)|\, dx_1 dy_2$$

so that

$$\|u\|_{L^2}^2 \le \|\nabla u\|_{L^1}^2,$$

which is a GNS inequality with constant $1$.

For higher dimensions ($N \ge 3$) and $p = 1$ see [Evans, Chapter 5], where Hölder's inequality is applied repeatedly in the course of the proof.

For $1 < p < N$, the case $p = 1$ is applied to a suitably chosen power of $|u|$; see [Evans, Chapter 5]. □

Now we adapt the GNS inequalities to bounded domains.

**Theorem 3.18** ($W_0^{1,p}, p < N$). *Suppose* $N = 1$. *If* $1 \le q \le \infty$, *then*

$$\|u\|_{L^q(U)} \le C\|u'\|_{L^1(U)}, \qquad u \in W_0^{1,1}(U),$$

*where* $C = C(q, U)$.

*Suppose* $N \ge 2$. *If* $1 \le p < N$ *and* $1 \le q \le p^*$, *then*

$$\|u\|_{L^q(U)} \le C\|\nabla u\|_{L^p(U)}, \qquad u \in W_0^{1,p}(U),$$

*where* $C = C(p, q, N, U)$.

*Proof.* Suppose $N \ge 1$. For $u$ smooth with compact support in $U$ we have

$$\begin{aligned} \|u\|_{L^q(U)} &\le C\|u\|_{L^{p^*}} \qquad \text{by Hölder (since } q \le p^* \text{ and } U \text{ has finite measure)} \\ &\le C\|\nabla u\|_{L^p(U)} \end{aligned}$$

by the GNS Theorem 3.17. Since smooth functions with compact support are dense in $W_0^{1,p}(U)$, the last inequality extends to $u \in W_0^{1,p}(U)$. (Some care is required since we do not know $u$ belongs to $L^q$; to prove that fact along with the rest of the result, take an approximating sequence and apply Fatou's lemma on the left side of the inequality.) □

***Exercise*** 3.15 ($W_0^{1,p}, p \ge N$). Show that if $p \in [N, \infty], q \in [1, \infty)$ then

$$\|u\|_{L^q(U)} \le C\|\nabla u\|_{L^p(U)}, \qquad u \in W_0^{1,p}(U),$$

where $C = C(p, q, N, U)$.

***Exercise*** 3.16. Show that if $p \in [1, \infty)$ then

$$\|u\|_{L^p(U)} \le C\|\nabla u\|_{L^p(U)}, \qquad u \in W_0^{1,p}(U),$$

where $C = C(p, N, U)$.

You already showed this last result for $p = 2$ in Exercise 3.13.

Theorem 3.18 cannot hold in general for functions that are not zero on the boundary — just take $u$ to be a constant function and notice the gradient on the right side of the theorem is zero while the $L^q$ norm on the left side is not. To handle such general Sobolev functions we change the right side of the inequality to use the full Sobolev norm, as follows.

**Theorem 3.19** ($W^{1,p}, p < N$)**.** *Suppose* $N = 1$*. If* $1 \le q \le \infty$ *then*

$$\|u\|_{L^q(U)} \le C\|u\|_{W^{1,p}(U)}, \qquad u \in W^{1,p}(U),$$

*where* $C = C(q, U)$*.*

*Suppose* $N \ge 2$*. If* $1 \le p < N$ *and* $1 \le q \le p^*$*, and* $\partial U$ *is* $C^1$*-smooth, then*

$$\|u\|_{L^q(U)} \le C\|u\|_{W^{1,p}(U)}, \qquad u \in W^{1,p}(U),$$

*where* $C = C(p, q, N, U)$*.*

The point of the result is again that Sobolev functions possess "better than expected" integrability.

*Proof.* Suppose $N = 1$, and write the domain as $U = (a, b)$. For any Lebesgue point $z \in [a, b]$, the fundamental theorem for Sobolev functions gives

$$|u(x)| \le |u(z)| + \Big| \int_z^x u'(y)\, dy \Big| \le |u(z)| + \|u'\|_{L^1(a,b)}.$$

Averaging over $z$ shows that $\|u\|_{L^\infty(a,b)} \le C\|u\|_{W^{1,1}(a,b)}$. Lastly, applying Hölder's inequality to the left and right sides completes the proof.

Suppose $N \ge 2$. Choose a larger domain $V$ that compactly contains $U$, and write $E : W^{1,p}(U) \to W_0^{1,p}(V)$ for the extension operator as in Section 3.6. Then

$$\begin{aligned} \|u\|_{L^q(U)} &\le \|Eu\|_{L^q(V)} && \text{since } Eu = u \text{ on } U \\ &\le C\|\nabla(Eu)\|_{L^p(V)} && \text{by Theorem 3.18} \\ &\le C\|Eu\|_{W^{1,p}(V)} && \text{by definition of the norms} \\ &\le C\|u\|_{W^{1,p}(U)} \end{aligned}$$

by the Extension Theorem 3.12. □

Next we examine the case $p > N$. Exercise 3.15 hints that $u$ might be bounded in that case. We prove the even stronger property of Hölder continuity, in the next theorem.

**Definition 3.20** (Hölder space)**.** Given $\gamma \in (0, 1]$ we define the **Hölder space** $C^{0,\gamma}(\overline{U})$ to consist of those functions for which the following **Hölder norm** is finite:

$$\|u\|_{C^{0,\gamma}(\overline{U})} = \sup_{x \in U} |u(x)| + \sup_{x \ne y} \frac{|u(x) - u(y)|}{|x - y|^\gamma}.$$

(To verify that this expression defines a norm is straightforward.)

Functions in the Hölder space are automatically bounded and continuous, and in fact Hölder continuous with exponent $\gamma$, meaning

$$|u(x) - u(y)| \leq C|x - y|^{\gamma}, \qquad x, y \in U,$$

for some constant $C$. Notice $u$ is uniformly continuous on $U$, and so it extends continuously to the closure $\overline{U}$.

***Exercise*** 3.17. Prove that the Hölder space $C^{0,\gamma}(\overline{U})$ is complete (and thus is a Banach space).

To consider the case $p > N$, we start with functions defined on all of $\mathbb{R}^N$.

**Theorem 3.21** (Morrey's inequality for $p > N$). *Let* $N < p \leq \infty$. *Then*

$$\|u\|_{C^{0,\gamma}(\mathbb{R}^N)} \leq C\|u\|_{W^{1,p}(\mathbb{R}^N)}, \qquad u \in C^1(\mathbb{R}^N),$$

*where* $C = C(p, N)$ *and*

$$\gamma = 1 - \frac{N}{p} \in (0, 1].$$

The moral here is that "big $p$ implies nice $u$".

*Proof.* Suppose $u \in C^1(\mathbb{R}^N)$ with $\|u\|_{W^{1,p}(\mathbb{R}^N)} < \infty$.
Step 1 — Average oscillation estimate. A constant $C = C(N)$ exists such that for all $x \in \mathbb{R}^N$ and $r > 0$:

$$\fint_{B(x,r)} |u(x) - u(z)| \, dz \leq Cr^{\gamma} \|\nabla u\|_{L^p(B(x,r))}. \tag{3.12}$$

The left side of (3.12) is the averaged oscillation of the function, near $x$, while the right side involves the derivative near $x$. We will assume $p < \infty$ in what follows. The case $p = \infty$ is similar.

To prove (3.12), first express $z \in B(x, r)$ as $z = x + sw$ where $s \in [0, r]$ and $w$ is a unit vector. Then by the Fundamental Theorem,

$$\begin{aligned}
|u(x) - u(z)| &= \Big| \int_0^s \frac{d}{dt} u(x + tw) \, dt \Big| \\
&= \Big| \int_0^s \nabla u(x + tw) \cdot w \, dt \Big| \\
&\leq \int_0^s |\nabla u(x + tw)||w| \, dt \\
&\leq \int_0^r |\nabla u(x + tw)| \, dt
\end{aligned}$$

since $s \le r$. Next, the left side of (3.12) is bounded by

$$
\begin{aligned}
&\frac{C}{r^N}\int_0^r \int_{\partial B(0,1)} |u(x)-u(x+sw)|\, dS(w) s^{N-1} ds \qquad \text{by spherical coordinates} \\
&\le \frac{C}{r^N}\int_0^r \int_{\partial B(0,1)} \int_0^r |\nabla u(x+tw)|\, dt\, dS(w) s^{N-1} ds \qquad \text{by above} \\
&= \frac{C}{r^N}\int_0^r s^{N-1} ds \int_0^r \int_{\partial B(0,1)} |\nabla u(x+tw)| t^{-(N-1)}\, dS(w) t^{N-1} dt \\
&= C \int_{B(x,r)} |\nabla u(y)||y-x|^{-(N-1)}\, dy
\end{aligned}
$$

where $y = x + tw, |y-x| = t$. Applying Hölder's inequality on the right side yields the $L^p$ norm of $\nabla u$ over $B(x,r)$ multiplied by the $L^{p'}$ norm of $|y|^{-(N-1)}$ over $B(0,r)$. By a change of variable, this last norm equals $r^\gamma$ times the $L^{p'}$ norm of $|y|^{-(N-1)}$ over the unit ball, where

$$
\gamma = -(N-1) + \frac{N}{p'} = -(N-1) + N\big(1 - \frac{1}{p}\big) = 1 - \frac{N}{p}.
$$

Note the $L^{p'}$ norm of $|y|^{-(N-1)}$ over the unit ball is finite, as we see by evaluating in spherical coordinates since

$$
-(N-1)p' + (N-1) = (N-1)(-\frac{p}{p-1} + 1) = -\frac{N-1}{p-1} > -1;
$$

in the last inequality we used that $p > N$.

Step 2 — Pointwise oscillation estimate. Given $x, y \in \mathbb{R}^N$, let $r = |x-y|$, take $V$ to be the ball of radius $r/2$ centered halfway between $x$ and $y$. Then

$$
\begin{aligned}
|u(x)-u(y)| &= \fint_V |u(x)-u(y)|\, dz \\
&\le \fint_V |u(x)-u(z)|\, dz + \fint_V |u(z)-u(y)|\, dz \\
&\le 2^N \fint_{B(x,r)} |u(x)-u(z)|\, dz + 2^N \fint_{B(y,r)} |u(z)-u(y)|\, dz
\end{aligned}
$$

since $V$ is contained in the balls $B(x,r)$ and $B(y,r)$ and has half their radius. Hence

$$
|u(x)-u(y)| \le C r^\gamma \|\nabla u\|_{L^p(\mathbb{R}^N)}
$$

by applying Step 1 with $x$ and again with $y$. That is,

$$\frac{|u(x) - u(y)|}{|x - y|^\gamma} \leq C\|\nabla u\|_{L^p(\mathbb{R}^N)}.$$

Step 3 — Size estimate. We have

$$\begin{aligned}
|u(x)| &= \fint_{B(x,1)} |u(x)|\, dz \\
&\leq \fint_{B(x,1)} |u(x) - u(z)|\, dz + \fint_{B(x,1)} |u(z)|\, dz \\
&\leq C\|\nabla u\|_{L^p(B(x,1))} + C\|u\|_{L^p(B(x,1))} \qquad \text{by Step 1 with } r = 1\text{, and Hölder,} \\
&\leq C\|u\|_{W^{1,p}(\mathbb{R}^N)}.
\end{aligned}$$

Step 4 — Complete the proof by combining Steps 2 and 3. □

Adapting the last result to a bounded domain gives:

**Theorem 3.22** ($p > N \implies u$ continuous). *Let* $N < p \leq \infty$ *and assume* $\partial U$ *is* $C^1$*-smooth. Each* $u \in W^{1,p}(U)$ *has a version* $u^* \in C^{0,\gamma}(\overline{U})$ *with*

$$\|u^*\|_{C^{0,\gamma}(\overline{U})} \leq C\|u\|_{W^{1,p}(U)},$$

*where* $\gamma = 1 - N/p \in (0, 1]$ *and* $C = C(p, N, U)$.

Recall $u^*$ is a **version** of $u$ if the two functions agree almost everywhere.

*Proof.* Following is a sketch of the proof. See [Evans] for details.

Assume $N < p < \infty$. We can reduce to considering $u \in C^1(\overline{U})$, by density. Take $u^* = u$. Extending $u$ to $Eu \in C^1_c(V)$ obviously increases the Hölder norm (since the domain has expanded). We can estimate the Hölder norm of $Eu$ with the Sobolev norm of $Eu$, by Morrey's Theorem 3.21. Lastly, the Extension Theorem 3.12 returns us to the Sobolev norm of $u$.

When $p = \infty$, the argument is more complicated since the density step is not valid; instead one uses a connection between Lipschitz functions and $W^{1,\infty}$. □

Collecting the recent theorems together yields better-than-expected integrability and Hölder continuity, in terms of higher Sobolev norms.

**Theorem 3.23** (Sobolev inequalities). *Assume* $\partial U$ *is* $C^1$*-smooth,* $k \geq 1$ *and* $1 \leq p < \infty$.

*(i) If*

$$kp < N$$

*then*

$$\|u\|_{L^q(U)} \leq C\|u\|_{W^{k,p}(U)}$$

*for* $u \in W^{k,p}(U)$*, where* $C = C(k, p, N, U)$ *and the exponent* $q$ *is determined from*

$$\frac{1}{q} = \frac{1}{p} - \frac{k}{N}.$$

*(ii) If*

$$kp > N$$

*then*

$$\|u\|_{C^{k-1-\lfloor N/p \rfloor, \gamma}(\overline{U})} \leq C\|u\|_{W^{k,p}(U)}$$

*for* $u \in W^{k,p}(U)$*, where* $C = C(k, p, N, U, \gamma)$ *and* $\gamma \in (0, 1)$ *satisfies*

$$\gamma = \begin{cases} \text{any positive number less than } 1 & \text{if } N/p \in \mathbb{Z}, \\ 1 + \lfloor N/p \rfloor - N/p & \text{if } N/p \notin \mathbb{Z}. \end{cases}$$

Note that part (ii) implicitly uses the continuous version of $u$.

*Proof.* For $k = 1$, simply combine Theorems 3.19 ($p < N$) and 3.22 ($p > N$).

For $k \geq 2$ one uses induction on $k$; see [Evans]. □

### Sobolev inequalities and smoothness

The Sobolev inequalities say that as $k$ and $p$ increase, the integrability or regularity of the function improves. In particular, part (ii) of the theorem implies that

$$kp > N \Longrightarrow u \in C(U),$$
$$kp > jp + N \Longrightarrow u \in C^j(U).$$

(The first claim is immediate, since part (ii) gives Hölder continuity of $u$. For the second claim, notice

$$\begin{aligned} kp > jp + N &\Longrightarrow k \geq j + \lfloor N/p \rfloor + 1 \\ &\Longrightarrow k - 1 - \lfloor N/p \rfloor \geq j \end{aligned}$$

so that $u \in C^j(U)$ by part (ii).) The theorem gives additional information too: it states the $j$th derivatives of $u$ are not only continuous but are Hölder continuous.

## Additional exercises

***Exercise*** 3.18 ($L^2$–$H^2$ interpolation).
(i) Prove that

$$\int_U |\nabla u|^2 \, dx \le C\Big(\int_U |u|^2 \, dx\Big)^{1/2}\Big(\int_U |D^2 u|^2 \, dx\Big)^{1/2}, \qquad u \in C_c^\infty(U),$$

where $D^2 u = \big[u_{x_j x_k}\big]_{j,k=1}^N$ denotes the Hessian matrix of $u$ and $|D^2u|^2 = \sum_{j,k}(u_{x_j x_k})^2$ is the sum of squares of its entries.
(ii) Show that the inequality in (i) holds for all $u \in H_0^2(U)$.
(iii) [Extra credit] Show that the inequality in (i) holds for all $u \in H^2 \cap H_0^1(U)$.
*Hint.* Start by extending Green's First Formula (3.1) to $u \in H_0^1$ and $v \in H^2$, by an approximation argument; there should be no boundary term.
*Notes.*
$H^0 := W^{0,2} = L^2, H^1 := W^{1,2}, H^2 := W^{2,2}$, and similarly $H_0^1 := W_0^{1,2}$.
The inequality proved in this problem is called an *interpolation inequality* because it says that if you have control over the function $u$ and its second derivatives, then you also have explicit control over the "inbetween" derivatives (the first derivatives). Many different interpolation inequalities are known, and we will use some of them later in the course when we prove regularity of solutions for PDEs.

***Exercise*** 3.19 (Partition of unity, part A). Suppose $U$ is a bounded open set in $\mathbb{R}^N$ with $U \Subset W$ for some open set $W$. Show that a function $\zeta \in C_c^\infty(W)$ exists such that $0 \le \zeta \le 1$ everywhere and $\zeta \equiv 1$ on $U$.

*Hint.* Construct a bounded open set $V$ with $U \Subset V \Subset W$ and mollify the indicator function $1_V$.

***Exercise*** 3.20 (Partition of unity, part B). Suppose $U \subset \mathbb{R}^N$ is a bounded open set with $U \Subset \cup_{i=1}^N W_i$ for some open sets $W_i$.

Show that functions $\zeta_i \in C_c^\infty(W_i)$ exists such that $0 \le \zeta_i \le 1$ everywhere and

$$\sum_{i=1}^N \zeta_i \equiv 1 \qquad \text{on } U.$$

*Hint.* First construct smooth functions $\eta_i \geq 0$ with $\sum_{i=1}^N \eta_i > 0$ on $\overline{U}$ and with $\eta_i$ supported in $W_i$. Then use $\eta_1, \dots, \eta_N$ to help construct $\zeta_1, \dots, \zeta_N$.

The collection of functions $\{\zeta_i\}$ is called a **partition of unity on** $U$ (subordinate to the covering $\{W_i\}$).

## 3.9 Compact imbedding of Sobolev spaces

**Definition 3.24** (Compact imbedding)**.** Let $X$ and $Y$ be Banach spaces with $X \subset Y$ (so that the two spaces have the same linear structure, although their norms might be different). We say $X$ is **imbedded** in $Y$, written $X \hookrightarrow Y$, if the imbedding map $\iota(x) = x$ is continuous:

$$\|x\|_Y \leq C\|x\|_X, \qquad x \in X,$$

for some constant $C$.

If in addition each bounded sequence in $X$ has a convergent subsequence in $Y$, then we say $X$ is **compactly imbedded** in $Y$, written $X \Subset Y$.

*Example* 3.25. Obviously $W^{1,p}(U) \hookrightarrow L^p(U)$, since the $L^p$ norm is bounded by the Sobolev norm. Further, $W^{1,p}(U) \hookrightarrow L^q(U)$ by the Sobolev inequalities in the previous section, when $p < N$ and $q \in [1, p^*]$.

The next theorems show that this imbedding is compact, except not in the endpoint case $q = p^*$.

Recall the domain $U$ is bounded by assumption, throughout this chapter.

**Theorem 3.26** (Rellich–Kondrachev for $W_0^{1,p}$)**.**
*(i) Suppose* $N = 1$. *If* $1 \leq q < \infty$ *then* $W_0^{1,1}(U) \Subset L^q(U)$.
*(ii) Suppose* $N \geq 2$. *If* $1 \leq p < N$ *and* $1 \leq q < p^*$ *then*

$$W_0^{1,p}(U) \Subset L^q(U).$$

**Theorem 3.27** (Rellich–Kondrachev for $W^{1,p}$)**.**
*(i) Suppose* $N = 1$. *If* $1 \leq q < \infty$ *then* $W^{1,1}(U) \Subset L^q(U)$.
*(ii) Suppose* $N \geq 2$. *If* $1 \leq p < N$ *and* $1 \leq q < p^*$ *and* $\partial U$ *is* $C^1$*-smooth, then*

$$W^{1,p}(U) \Subset L^q(U).$$

Intuitively, the theorems say that by giving up a derivative we gain compactness.

*Proof of Theorem 3.26.* The hypotheses ensure $1 \leq q < p^*$ when $N \geq 2$, and the same holds when $N = 1$ since then $p = 1$ and $p^* = \infty$ (by definition).

Step 1 — $L^q$-mollification rate. Since

$$\frac{1}{p^*} < \frac{1}{q} \leq 1,$$

there exists $\theta \in (0, 1]$ such that

$$\frac{1}{q} = \frac{1-\theta}{p^*} + \frac{\theta}{1}.$$

(Notice $\theta > 0$ because $q < p^*$. That is why we exclude the endpoint case $q = p^*$ in the theorem.) We claim a constant $C$ exists, independent of $u$, such that

$$\|u^\varepsilon - u\|_{L^q(U)} \leq C\varepsilon^\theta \|u\|_{W^{1,p}(U)}, \qquad \forall u \in C_c^\infty(U), \tag{3.13}$$

whenever $\varepsilon > 0$. (That is, mollification converges at a rate determined by the $W^{1,p}$-norm of the function.) To prove this claim, we recall our standard mollifier $\eta$ and compute that

$$\begin{aligned}
u^\varepsilon(x) - u(x) &= (\eta_\varepsilon * u)(x) - u(x) \\
&= \int_{B(0,1)} \eta(y)\big(u(x - \varepsilon y) - u(x)\big)\, dy \qquad \text{since } \int_{B(0,1)} \eta(y)\, dy = 1 \\
&= \int_{B(0,1)} \eta(y) \int_0^1 \frac{d}{dt} u(x - \varepsilon t y)\, dt dy \\
&= -\varepsilon \int_{B(0,1)} \eta(y) \int_0^1 (\nabla u)(x - \varepsilon t y) \cdot y\, dt dy.
\end{aligned}$$

Integrating with respect to $x \in U$, we find

$$\int_U |u^\varepsilon(x) - u(x)|\, dx \leq \varepsilon \int_{B(0,1)} \eta(y) \int_0^1 \int_U \big|\nabla u(x - \varepsilon t y)\big|\, dx dt dy$$

since $|y| \leq 1$. The inner integral is

$$\int_U \big|\nabla u(x - \varepsilon t y)\big|\, dx \leq \int_{\mathbb{R}^N} |\nabla u|\, dx = \int_U |\nabla u|\, dx \leq C\|\nabla u\|_{L^p(U)} \leq C\|u\|_{W^{1,p}(U)},$$

and so

$$\int_U |u^\varepsilon(x) - u(x)|\, dx \le C\varepsilon \|u\|_{W^{1,p}(U)}. \tag{3.14}$$

By interpolation (Exercise 3.22 below) and (3.14),

$$\begin{aligned}\|u^\varepsilon - u\|_{L^q(U)} &\le \|u^\varepsilon - u\|_{L^1(U)}^{\theta} \|u^\varepsilon - u\|_{L^{p^*}(U)}^{1-\theta} \\ &\le \left(C\varepsilon\|u\|_{W^{1,p}(U)}\right)^{\theta} \left(\|u^\varepsilon\|_{L^{p^*}(U)} + \|u\|_{L^{p^*}(U)}\right)^{1-\theta}.\end{aligned}$$

For the final term, notice

$$\|u^\varepsilon\|_{L^{p*}(U)} \le \|\eta_\varepsilon * u\|_{L^{p*}(\mathbb{R}^N)} \le \|\eta_\varepsilon\|_{L^1} \|u\|_{L^{p*}(\mathbb{R}^N)} = \|u\|_{L^{p*}(U)},$$

and $\|u\|_{L^{p*}(U)} \le C\|u\|_{W^{1,p}(U)}$ by the Sobolev inequality. So we get (3.13).

Step 2 — Reduction to smooth functions with compact support. Suppose $\{v_j\}$ is a bounded sequence in $W_0^{1,p}(U)$. To prove the theorem, we want to extract a subsequence that converges in $L^q$. By density we may choose functions $\widetilde{v}_j \in C_c^\infty(U)$ such that $\|v_j - \widetilde{v}_j\|_{W^{1,p}(U)} \to 0$ as $j \to \infty$. Hence $\|v_j - \widetilde{v}_j\|_{L^q(U)} \to 0$ by the GNS Theorem 3.18. Thus if we construct a subsequence of $\{\widetilde{v}_j\}$ that converges in $L^q$, then the corresponding subsequence of $\{v_j\}$ must also converge in $L^q$. Therefore we may suppose from now on that $v_j \in C_c^\infty(U)$.

Step 3 — Almost Cauchy subsequence. We will construct for each $\delta > 0$ an "almost Cauchy in $L^q$" subsequence $\{v_{j_k}\}$, satisfying

$$\limsup_{k,\ell\to\infty} \|v_{j_k} - v_{j_\ell}\|_{L^q(U)} \le \delta.$$

First, since the sequence $\{v_j\}$ is bounded we can define $M = \sup_j \|v_j\|_{W^{1,p}(U)} < \infty$. Choose $\varepsilon > 0$ such that $CM\varepsilon^\theta < \delta/2$ (using here that $\theta > 0$). Then by Step 1,

$$\|v_j^\varepsilon - v_j\|_{L^q(U)} < \frac{\delta}{2}, \qquad j \ge 1.$$

Thus it suffices to find a subsequence such that $\{v_{j_k}^\varepsilon\}$ converges in $L^q(U)$ as $k \to \infty$. Such a subsequence exists by the Arzelà–Ascoli compactness theorem: simply check that the functions $v_j^\varepsilon$ (with $\varepsilon$ fixed and $j \ge 1$) are uniformly bounded since

$$|v_j^\varepsilon(x)| \le \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)|v_j(y)|\, dy \le \frac{C}{\varepsilon^N}\|v_j\|_{L^1(U)} \le \frac{CM}{\varepsilon^N},$$

and are uniformly equicontinuous because

$$|\nabla v_j^\varepsilon(x)| \le \int_{B(x,\varepsilon)} |\nabla \eta_\varepsilon(x-y)|\,|v_j(y)|\,dy \le \frac{C}{\varepsilon^{N+1}}\|v_j\|_{L^1(U)} \le \frac{CM}{\varepsilon^{N+1}}.$$

Step 4 — Diagonal argument. Use Step 3 with $\delta = 1/2, 1/3, 1/4, \ldots$ to extract successive subsequences, and then take the diagonal subsequence. Denoting it by $\{v_n\}$, we see from Step 2 that

$$\limsup_{m,n\to\infty} \|v_m - v_n\|_{L^q(U)} = 0.$$

That is, we have found a Cauchy subsequence. This subsequence converges in $L^q(U)$, which proves compactness of the imbedding of $W_0^{1,p}$ into $L^q$. □

*Proof of Theorem 3.27.* We know $W^{1,p}(U)$ imbeds into $L^q(U)$, by Theorem 3.19. To show the imbedding is compact, we let $V$ be a bounded domain that compactly contains $U$, and observe that:

$\{u_k\}$ bounded in $W^{1,p}(U)$

$\Longrightarrow \{Eu_k\}$ bounded in $W_0^{1,p}(V)$ by the Extension Theorem 3.12

$\Longrightarrow \{Eu_{k_j}\}$ converges in $L^q(V)$ by the compact imbedding in Theorem 3.26

$\Longrightarrow \{u_{k_j}\}$ converges in $L^q(U)$, by restricting to $U$.

Hence $W^{1,p}(U)$ imbeds compactly into $L^q(U)$. □

**Corollary 3.28.** *Let* $1 \le p \le \infty$.
*(i) Then*

$$W_0^{1,p}(U) \Subset L^p(U).$$

*(ii) If* $\partial U$ *is* $C^1$*-smooth then*

$$W^{1,p}(U) \Subset L^p(U).$$

***Exercise*** 3.21. Prove Corollary 3.28. *Hint.* When $N > 1$, consider the three cases $p < N, N \le p < \infty, p = \infty$. When $p = \infty$, your replacement for Rellich–Kondrachev will be the Arzelà–Ascoli pre-compactness theorem.

***Exercise*** 3.22 ($L^p$ interpolation). Let $(X, \mu)$ be a measure space and suppose $1 \le r \le s \le t \le \infty$. Choose $\theta \in [0,1]$ such that

$$\frac{1}{s} = \frac{1-\theta}{t} + \frac{\theta}{r}.$$

If $f \in L^r \cap L^t(\mu)$ then $f \in L^s(\mu)$ with norm

$$\|f\|_s \leq \|f\|_r^\theta \|f\|_t^{1-\theta}.$$

*Hint.* $|f| = |f|^\theta |f|^{1-\theta}$. Apply Hölder's inequality.

***Exercise*** 3.23. Find functions $u_n \in W^{1,2}(0,1)$ such that the sequence $\{u_n\}$ is bounded but non-convergent in $W^{1,2}(0,1)$, and yet $u_n \to 0$ in $L^2(0,1)$. (*Hint.* Suppose the derivatives $u_n'$ are orthonormal in $L^2$. See Exercise 2.22 parts (ii) or (iii).) Similar examples can then be constructed in higher dimensions, to illustrate the conclusion of the Rellich–Kondrachev theorems.

***Exercise*** 3.24. Show the imbedding of $W_0^{1,p}(U)$ into $L^q(U)$ is *not* compact in the endpoint case ($q = p^*$). *Scaling hint.* Assume $U$ contains the origin, and consider the family of functions $v_\lambda(x) = \lambda^{-1+N/p} u(\lambda x)$ as $\lambda \to \infty$, for some fixed $u \in C_c^\infty(\mathbb{R}^N)$.

Our first application of compact imbedding is to control deviations from the mean value in terms of the gradient norm. Denote the mean value of $u$ by

$$m = \frac{1}{|U|} \int_U u \, dx.$$

**Theorem 3.29** (Poincaré inequality). *Assume $U$ is connected and $\partial U$ is $C^1$-smooth. If $1 \leq p \leq \infty$ then*

$$\|u - m\|_{L^p(U)} \leq C \|\nabla u\|_{L^p(U)}, \qquad u \in W^{1,p}(U),$$

*where $C = C(N, p, U)$.*

*Proof.* Suppose no such constant $C$ exists. Then for each $k \geq 1$ there exists some $u_k \in W^{1,p}(U)$ for which

$$\|u_k - m_k\|_{L^p} > k \|\nabla u_k\|_{L^p}.$$

We normalize the function by letting

$$v_k = \frac{u_k - m_k}{\|u_k - m_k\|_{L^p}}.$$

Then $\int_U v_k \, dx = 0$, $\|v_k\|_{L^p} = 1$ and $\|\nabla v_k\|_{L^p} < 1/k \to 0$.

In particular, $\{v_k\}$ is bounded in $W^{1,p}(U)$. By the Rellich–Kondrachev Corollary 3.28(ii), after passing to a subsequence we may assume $v_k \to v$ in

$L^p$, for some $v \in L^p(U)$. Hence $\|v\|_{L^p} = 1$, and also $\int_U v \, dx = \lim \int_U v_k \, dx = 0$.

On the other hand, $\nabla v = 0$ weakly because for each test function $\phi \in C_c^\infty(U)$ and each $j$, we have

$$\begin{aligned}\int_U v\phi_{x_j} \, dx &= \lim_k \int_U v_k \phi_{x_j} \, dx \qquad \text{since } v_k \to v \text{ in } L^p \\ &= -\lim_k \int_U (v_k)_{x_j} \phi \, dx \\ &= 0,\end{aligned}$$

using here that $\nabla v_k \to 0$ in $L^p$. Finally, since $\nabla v = 0$ weakly and $U$ is connected, we conclude $v = c$ a.e. by Exercise 3.12. This constant $c$ must be $0$ because $v$ has mean value zero. Thus $v = 0$ a.e., contradicting that $\|v\|_{L^p} = 1$. □

***Exercise*** 3.25 (Poincaré on balls). Show that if the domain in Theorem 3.29 is a ball of radius $r$, then the Poincaré inequality holds with a constant $C = C(N, p)r$ that depends linearly on $r$.

***Exercise*** 3.26. Let $X$ and $Y$ be Hilbert spaces with $X \hookrightarrow Y$. Prove that if $u_j \rightharpoonup u$ weakly in $X$ then $u_j \rightharpoonup u$ weakly in $Y$.
*Hint.* Consider a bounded linear functional $F$ on $Y$ and show it is also a bounded linear functional on $X$.

For example, this exercise shows that weak convergence in $H_0^1$ implies weak convergence in $L^2$.

## 3.10 *Application:* Poisson's equation via Calculus of Variations (Dirichlet Principle)

Define a nonlinear "energy" functional $F : H_0^1(U) \to \mathbb{R}$ by

$$F(u) = \int_U \Big(\frac{1}{2}|\nabla u|^2 - fu\Big) \, dx,$$

where $f \in L^2(U)$ is some given function. We will show that $F$ attains a minimum value, and that the minimizing $u$ satisfies Poisson's equation $-\Delta u = f$ weakly.

This method is more difficult than the solution using Riesz Representation that we gave in Section 2.2 and Exercise 3.14. It has the advantage, though, of suggesting how to handle certain nonlinear PDEs in the field of calculus of variations.

The next lemma controls the Sobolev norm in terms of the energy.

**Lemma 3.30** (Lower bound on the energy).

$$F(u) \geq C_1 \|u\|^2_{H^1(U)} - C_2 \|f\|^2_{L^2(U)}, \qquad u \in H^1_0(U),$$

*for some constants* $C_1, C_2 > 0$ *that depend on* $N$ *and* $U$.

*Proof.* The elementary "**Cauchy with epsilon**" inequality

$$ab \leq \frac{1}{2\varepsilon} a^2 + \frac{\varepsilon}{2} b^2 \tag{3.15}$$

follows from completing the square, for all $a, b \in \mathbb{R}$ and $\varepsilon > 0$. Using this inequality we find

$$\begin{aligned} \Big| \int_U fu \, dx \Big| &\leq \int_U \big( \frac{1}{2\varepsilon} |f|^2 + \frac{\varepsilon}{2} |u|^2 \big) \, dx \\ &= \frac{1}{2\varepsilon} \|f\|^2_{L^2} + \frac{\varepsilon}{2} \|u\|^2_{L^2} \\ &= \frac{1}{2\varepsilon} \|f\|^2_{L^2} - \frac{\varepsilon}{2} \|u\|^2_{L^2} + \varepsilon \|u\|^2_{L^2} \\ &\leq \frac{1}{2\varepsilon} \|f\|^2_{L^2} - \frac{\varepsilon}{2} \|u\|^2_{L^2} + \varepsilon C \|\nabla u\|^2_{L^2} \end{aligned}$$

by Exercise 3.16 (consequence of the Sobolev inequalities). Substituting into the definition of $F(u)$ gives that

$$F(u) \geq \big( \frac{1}{2} - \varepsilon C \big) \|\nabla u\|^2_{L^2} + \frac{\varepsilon}{2} \|u\|^2_{L^2} - \frac{1}{2\varepsilon} \|f\|^2_{L^2}.$$

Choosing $\varepsilon$ small enough that $1/2 - \varepsilon C \geq \varepsilon/2$ now proves the lemma for $C_1 = \varepsilon/2$ and $C_2 = 1/2\varepsilon$. □

We write the lowest energy of $F$ as

$$I = \inf\{F(u) : u \in H^1_0(U)\}.$$

We will use weak compactness and the compact imbedding of $H^1_0 \hookrightarrow L^2$ to show the infimum is attained.

**Proposition 3.31** (Existence of an energy minimizer)**.** $I = F(u)$ *for some* $u \in H^1_0(U)$.

*Proof.*
Step 1 — Use compact imbedding. Choose an infimizing sequence $u_k$ such that $F(u_k) \searrow I$. The sequence is bounded in $H^1_0(U)$ by Lemma 3.30, since $F(u_k)$ is bounded. Hence by the Rellich–Kondrachev Corollary 3.28(i) with $p = 2$, after passing to a subsequence we can assume

$$u_k \to u \qquad \text{in } L^2,$$

for some $u \in L^2(U)$. (We do not know yet whether $u$ belongs to $H^1_0$.) Thus in particular $u_k \rightharpoonup u$ weakly in $L^2$.
Step 2 — Use weak sequential compactness of the closed ball. Theorem 2.31 implies (after passing to a subsequence) that

$$u_k \rightharpoonup \widetilde{u} \qquad \text{weakly in } H^1_0$$

for some $\widetilde{u} \in H^1_0(U)$. Hence $u_k \rightharpoonup \widetilde{u}$ weakly in $L^2$, by Exercise 3.26. The uniqueness of weak limits implies now that $\widetilde{u} = u$.
Step 3 — Combine weak and norm information. Since $u_k \rightharpoonup u$ weakly in $H^1_0$ we know

$$\|u\|_{H^1} \leq \liminf \|u_k\|_{H^1}$$

by Exercise 2.20. Also $\|u_k\|_{L^2} \to \|u\|_{L^2}$ and $\int_U f u_k \, dx \to \int_U f u \, dx$ because $u_k \to u$ in $L^2$. Therefore

$$F(u) \leq \liminf F(u_k) = I.$$

In the other direction, $F(u) \geq I$ since $I$ is the infimum of $F$. Hence $F(u) = I$. □

Now that we have existence of a minimizer, we can study its properties.

**Theorem 3.32** (Weak solution of Poisson's equation with Dirichlet boundary condition)**.** *The energy-minimizing* $u \in H^1_0(U)$ *constructed in Proposition 3.31 solves*

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

*weakly, and this weak solution is unique.*

*Proof.* We will derive the "Euler–Lagrange" condition for the energy minimizer $u$. The idea is simply to vary $u$ in some direction $v$ and invoke the first derivative test from calculus.

So consider $v \in H^1_0(U)$ and let $g(t) = F(u + tv)$ for $t \in \mathbb{R}$. Then $g$ is a real valued function whose minimum occurs at $t = 0$. Hence

$$\begin{aligned} 0 &= g'(0) \\ &= \frac{d}{dt} \int_U \big(\frac{1}{2}|\nabla u + t\nabla v|^2 - fu - tfv\big)\, dx\Big|_{t=0} \\ &= \int_U \big(\nabla u \cdot \nabla v - fv\big)\, dx, \end{aligned}$$

which means

$$\int_U \nabla u \cdot \nabla v \, dx = \int_U fv \, dx, \qquad \forall v \in H^1_0(U).$$

This last equation is the weak form of Poisson's equation with Dirichlet boundary condition.

For uniqueness, suppose two solutions $u_1$ and $u_2$ exist. Then $\Delta(u_1 - u_2) = 0$ weakly, so that

$$\int_U \nabla(u_1 - u_2) \cdot \nabla v \, dx = 0$$

for all $v \in H^1_0(U)$. Choosing $v = u_1 - u_2$ shows $\|\nabla(u_1 - u_2)\|_{L^2} = 0$, and hence $u_1 - u_2 = 0$ by Exercise 3.13. □

# Chapter 4

# Discrete spectral theory for symmetric elliptic operators

## 4.1 Abstract spectral theory for sesquilinear forms

**Reference** [BlanchardBrüning] Chapter 6

We aim to use Sobolev space theory to solve selfadjoint elliptic PDEs on bounded domains. To achieve these goals we will use an ONB of eigenfunctions. We begin with an abstract approach to spectral theory from functional analysis, and then apply it to obtain weak eigenfunctions of partial differential operators.

### PDE preview — weak eigenfunctions

Consider the eigenfunction equation $-\Delta u = \lambda u$ for the Laplacian, in a domain $U$. Multiply by $v \in H_0^1(U)$ and integrate formally by parts to obtain

$$\int_U \nabla u \cdot \nabla v \, dx = \lambda \langle u, v \rangle_{L^2(U)}, \qquad v \in H_0^1(U).$$

We call this condition the "weak form" of the eigenfunction equation. To prove existence of ONBs of such weak eigenfunctions, we generalize to a Hilbert space setting.

**Hypotheses**

Consider infinite dimensional Hilbert spaces $\mathcal{H}$ and $\mathcal{K}$ over $\mathbb{R}$ (or $\mathbb{C}$).

$\mathcal{H}$: inner product $\langle u, v\rangle_{\mathcal{H}}$, norm $\|u\|_{\mathcal{H}}$

$\mathcal{K}$: inner product $\langle u, v\rangle_{\mathcal{K}}$, norm $\|u\|_{\mathcal{K}}$

Assume $\mathcal{H}$ is separable and:

1. $\mathcal{K}$ is continuously and densely imbedded in $\mathcal{H}$, meaning there exists a continuous linear injection $\iota : \mathcal{K} \to \mathcal{H}$ with $\iota(\mathcal{K})$ dense in $\mathcal{H}$.

2. The imbedding $\mathcal{K} \hookrightarrow \mathcal{H}$ is **compact**, meaning if $B$ is a bounded subset of $\mathcal{K}$ then $B$ is precompact when considered as a subset of $\mathcal{H}$. (Equivalently, every bounded sequence in $\mathcal{K}$ has a subsequence that converges in $\mathcal{H}$.)

3. We have a map $a : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ (or $\mathbb{C}$) that is sesquilinear, continuous, and **symmetric**, meaning

$$\begin{aligned} u &\mapsto a(u, v) \text{ is linear, for each fixed } v, \\ v &\mapsto a(u, v) \text{ is linear (or conjugate linear), for each fixed } u, \\ |a(u, v)| &\leq (\text{const.})\|u\|_{\mathcal{K}}\|v\|_{\mathcal{K}} \qquad \text{(boundedness)} \\ a(v, u) &= a(u, v) \quad (\text{or } = \overline{a(u, v)}) \end{aligned}$$

4. $a$ is **elliptic** on $\mathcal{K}$, meaning

$$a(u, u) \geq c\|u\|_{\mathcal{K}}^2 \qquad \forall u \in \mathcal{K},$$

for some $c > 0$.

Boundedness and ellipticity imply $a(u, u) \asymp \|u\|_{\mathcal{K}}^2$, and since $a$ is symmetric:

$a(u, v)$ is an inner product equivalent to $\langle u, v\rangle_{\mathcal{K}}$.

*Example* 4.1. The key example for us is when $\mathcal{H} = L^2(U), \mathcal{K} = H_0^1(U), a(u, v) = \int_U \nabla u \cdot \nabla v \, dx$. Compactness of the imbedding $H_0^1 \Subset L^2$ follows from Rellich–Kondrachev Corollary 3.28(i), while density is obvious since $C_c^\infty$ is dense in $L^2$. Ellipticity of $a$ follows from the Sobolev inequalities (or simply use Exercise 3.13).

**Theorem 4.2** (Spectral theorem)**.** *Under the hypotheses above, there exist vectors* $u_1, u_2, u_3, \ldots \in \mathcal{K}$ *and numbers*

$$0 < \gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \cdots \to \infty$$

*such that:*

- $u_j$ *is an eigenvector of* $a(\cdot,\cdot)$ *with eigenvalue* $\gamma_j$*, meaning*

$$a(u_j, v) = \gamma_j \langle u_j, v\rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{K}, \tag{4.1}$$

- $\{u_j\}$ *is an ONB for* $\mathcal{H}$*,*
- $\{u_j/\sqrt{\gamma_j}\}$ *is an ONB for* $\mathcal{K}$ *with respect to the* $a$*-inner product.*

*The series decomposition*

$$f = \sum_j \langle f, u_j\rangle_{\mathcal{H}}\, u_j \tag{4.2}$$

*converges in* $\mathcal{H}$ *for each* $f \in \mathcal{H}$*, and converges in* $\mathcal{K}$ *for each* $f \in \mathcal{K}$*.*

The idea is to show that a certain "inverse" operator associated with $a$ is compact and selfadjoint on $\mathcal{H}$. This approach makes sense in terms of differential equations, where $a$ would correspond to a differential operator such as $-\Delta$ (which is unbounded) and the inverse would correspond to an integral operator $(-\Delta)^{-1}$ (which is bounded, and in fact compact, on suitable domains). Indeed, we will begin by solving the analogue of $-\Delta u = f$ weakly in our Hilbert space setting, with the help of the Riesz Representation Theorem. This part of the proof parallels our solution of the Poisson equation in Section 2.2.

*Proof of Theorem 4.2.* Step 1 — the solution map. We claim that for each $f \in \mathcal{H}$ there exists a unique $u \in \mathcal{K}$ such that

$$a(u, v) = \langle f, v\rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{K}, \tag{4.3}$$

and that the solution map

$$\begin{aligned} B : &\mathcal{H} \to \mathcal{K} \\ &f \mapsto u \end{aligned}$$

is linear and bounded. To prove this claim, fix $f \in \mathcal{H}$ and define a bounded linear functional $F(v) = \langle v, f\rangle_{\mathcal{H}}$ on $\mathcal{K}$, noting for the boundedness that

$$\begin{aligned} |F(v)| &\leq \|v\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \\ &\leq (\text{const.}) \|v\|_{\mathcal{K}} \|f\|_{\mathcal{H}} \qquad \text{since } \mathcal{K} \text{ is imbedded in } \mathcal{H} \\ &\leq (\text{const.}) a(v, v)^{1/2} \|f\|_{\mathcal{H}} \end{aligned}$$

by ellipticity. Hence by the Riesz Representation Theorem on $\mathcal{K}$ (with respect to the $a$-inner product and norm on $\mathcal{K}$), there exists a unique $u \in \mathcal{K}$ such that $F(v) = a(v, u)$ for all $v \in \mathcal{K}$. That is,

$$\langle v, f \rangle_{\mathcal{H}} = a(v, u) \qquad \forall v \in \mathcal{K},$$

as desired for (4.3). Thus the map $B : f \mapsto u$ is well defined. Clearly it is linear. And

$$a(u, u) = |F(u)| \leq (\text{const.}) a(u, u)^{1/2} \|f\|_{\mathcal{H}}.$$

Hence $a(u, u)^{1/2} \leq (\text{const.}) \|f\|_{\mathcal{H}}$, so that $B$ is bounded from $\mathcal{H}$ to $\mathcal{K}$, which proves our initial claim.

For later use, note that (4.3) can be rephrased as saying

$$a(Bf, v) = \langle f, v \rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{K}. \tag{4.4}$$

Step 2 — $B$ is compact and selfadjoint. Note $B : \mathcal{H} \to \mathcal{K} \to \mathcal{H}$ is compact, since $\mathcal{K}$ imbeds compactly into $\mathcal{H}$. Further, $B$ is selfadjoint on $\mathcal{H}$, since for all $f, g \in \mathcal{H}$ we have

$$\begin{aligned} \langle Bf, g \rangle_{\mathcal{H}} &= \overline{\langle g, Bf \rangle_{\mathcal{H}}} \\ &= \overline{a(Bg, Bf)} && \text{by (4.4)}, \\ &= a(Bf, Bg) && \text{by symmetry of } a, \\ &= \langle f, Bg \rangle_{\mathcal{H}} && \text{by (4.4)}, \end{aligned}$$

which implies that $B^* = B$.

Step 3 — apply the spectral theorem for compact, selfadjoint operators. Spectral Theorem 4.6 below (which uses the separability assumption on $\mathcal{H}$) provides an ONB for $\mathcal{H}$ consisting of eigenvectors of $B$. The eigenvalues of $B$ are nonzero because $B$ is injective: $Bf = 0$ would imply $\langle f, v \rangle_{\mathcal{H}} = 0$ for all $v \in \mathcal{K}$ by (4.4), so that $f = 0$ by density of $\mathcal{K}$ in $\mathcal{H}$. Thus the spectral theorem says

$$Bu_j = \beta_j u_j$$

for some ONB of eigenvectors $\{u_j\}$ with nonzero real eigenvalues $\beta_j \to 0$. After reordering, we may assume the eigenvalues decrease in magnitude: $|\beta_1| \geq |\beta_2| \geq \cdots \to 0$.

The decomposition (4.2) holds in $\mathcal{H}$ because $\{u_j\}$ forms an ONB for $\mathcal{H}$.

Since $\beta_j \neq 0$ we may divide to obtain $u_j = B(u_j/\beta_j)$. Thus $u_j$ belongs to the range of $B$, and so $u_j \in \mathcal{K}$.

The eigenvalues are all positive, since

$$\beta_j a(u_j, v) = a(Bu_j, v) = \langle u_j, v\rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{K}$$

and choosing $v = u_j \in \mathcal{K}$ and using ellipticity shows that $\beta_j > 0$. Thus we see that the reciprocal numbers $0 < \gamma_j \stackrel{\text{def}}{=} 1/\beta_j \to \infty$ satisfy

$$a(u_j, v) = \gamma_j \langle u_j, v\rangle_{\mathcal{H}} \qquad \forall v \in \mathcal{K},$$

which is (4.1). (*Remark.* This step of taking reciprocal eigenvalues corresponds intuitively to inverting the solution map $B$ to obtain a "direct" map $A : u \mapsto f$.)

Finally, we have $a$-orthonormality of the set $\{u_j/\sqrt{\gamma_j}\}$:

$$\begin{aligned} a(u_j, u_k) &= \gamma_j \langle u_j, u_k\rangle_{\mathcal{H}} \\ &= \gamma_j \delta_{jk} \\ &= \sqrt{\gamma_j}\sqrt{\gamma_k}\,\delta_{jk}. \end{aligned}$$

This orthonormal set is complete in $\mathcal{K}$, because if $v \in \mathcal{K}$ has $a(u_j, v) = 0$ for all $j$, then $\langle u_j, v\rangle_{\mathcal{H}} = 0$ for all $j$ by (4.1), so that $v = 0$ in $\mathcal{H}$ and hence in $\mathcal{K}$ (because the imbedding $\iota : \mathcal{K} \to \mathcal{H}$ is injective). Therefore each $f \in \mathcal{K}$ can be decomposed as

$$f = \sum_j a(f, u_j/\sqrt{\gamma_j})\, u_j/\sqrt{\gamma_j}$$

with convergence in $\mathcal{K}$, and this decomposition reduces to (4.2) because $a(f, u_j) = \gamma_j \langle f, u_j\rangle_{\mathcal{H}}$. □

*Remark* 4.3. Eigenvectors corresponding to distinct eigenvalues are automatically orthogonal, since

$$\begin{aligned} (\gamma_j - \gamma_k)\langle u_j, u_k\rangle_{\mathcal{H}} &= \gamma_j \langle u_j, u_k\rangle_{\mathcal{H}} - \overline{\gamma_k \langle u_k, u_j\rangle_{\mathcal{H}}} \\ &= a(u_j, u_k) - \overline{a(u_k, u_j)} \\ &= 0 \end{aligned}$$

by symmetry of $a$.

## 4.2 Spectral theorem for compact, selfadjoint operators

In the previous section we needed the existence of an ONB of eigenvectors of a compact selfadjoint operator. In this section we prove the needed theorem.

Let $\mathsf{H}$ be a Hilbert space. We begin with two useful lemmas about weakly convergent sequences.

**Lemma 4.4** (Continuous linear maps preserve weak convergence)**.** *Suppose* $T : H \to H$ *is linear and bounded. If* $u_k \rightharpoonup u$ *weakly then* $Tu_k \rightharpoonup Tu$ *weakly.*

*Proof.* For all $v \in H$ we have

$$\begin{aligned} \langle Tu_k, v \rangle &= \langle u_k, T^* v \rangle && \text{where } T^* \text{ is the Hilbert space adjoint of } T \\ &\to \langle u, T^* v \rangle && \text{since } u_k \rightharpoonup u_k \text{ weakly} \\ &= \langle Tu, v \rangle. \end{aligned}$$

□

**Lemma 4.5** (Compact linear maps take weak convergence to norm convergence)**.** *Suppose* $T : H \to H$ *is linear, bounded and compact. If* $u_k \rightharpoonup u$ *weakly then* $Tu_k \to Tu$ *in norm. Also* $\langle u_k, Tu_k \rangle \to \langle u, Tu \rangle$*.*

*Proof.* Consider an arbitrary subsequence $\{u_{k_\ell}\}$, which for notational simplicity we write as $\{v_\ell\}$. Since $Tu_k \rightharpoonup Tu$ weakly by Lemma 4.4, in particular we have $Tv_\ell \rightharpoonup Tu$ weakly.

The weakly convergent sequence $\{u_k\}$ is bounded, by Proposition 2.33. Compactness of $T$ therefore implies norm convergence of some subsequence of $\{Tv_\ell\}$, say $Tv_{\ell_m} \to v$ for some $v \in H$. This norm convergence implies also the weak convergence $Tv_{\ell_m} \rightharpoonup v$.

Combining the last two paragraphs shows that $Tu = v$. Thus $Tv_{\ell_m} \to Tu$ in norm. We have shown that each subsequence of $\{u_k\}$ has a further subsequence whose image under $T$ converges to $Tu$, and so $Tu_k \to Tu$.

Further,

$$\langle u_k, Tu_k \rangle - \langle u, Tu \rangle = \langle u_k, Tu_k - Tu \rangle + \langle u_k - u, Tu \rangle.$$

The first inner product on the right side converges to $0$ because $\{u_k\}$ is bounded and $Tu_k - Tu \to 0$. The second inner product converges to $0$ because $u_k \rightharpoonup u$ weakly. □

Now comes the main result of the section, giving an ONB of eigenvectors.

**Theorem 4.6** (Spectral theorem for compact selfadjoint operators)**.** *Assume* $H$ *is a separable, infinite dimensional Hilbert space, and* $B : H \to H$ *is a linear, compact, selfadjoint operator.*

*Then* $H$ *has a countable ONB* $\{u_k\}$ *consisting of eigenvectors of* $B$*, say*

$$Bu_k = \beta_k u_k$$

*for some eigenvalues* $\beta_k \in \mathbb{R}$.

*Further, the set* $\{\beta_k : |\beta_k| > \varepsilon\}$ *is finite for each* $\varepsilon > 0$*, and so if* $B$ *has infinitely many nonzero eigenvalues then they can be arranged as a sequence converging to* $0$*. In particular, each nonzero eigenvalue has finite multiplicity (meaning if* $\beta \neq 0$ *is an eigenvalue then the corresponding eigenspace is finite dimensional).*

The finite dimensional version of the theorem simply says that a selfadjoint matrix (either real symmetric or complex Hermitian) possesses an ONB of eigenvectors.

*Proof.* Note that $\langle u, Bu\rangle$ is real for each $u \in H$, because by selfadjointness of $B$ we have

$$\langle u, Bu\rangle = \langle u, B^*u\rangle = \langle Bu, u\rangle = \overline{\langle u, Bu\rangle}.$$

Define

$$m = \inf_{u\neq 0} \frac{\langle u, Bu\rangle}{\langle u, u\rangle}, \qquad M = \sup_{u\neq 0} \frac{\langle u, Bu\rangle}{\langle u, u\rangle},$$

so that $-\infty < m \leq M < \infty$.

Step 1 — Finding the largest eigenvalue. Assume $M > 0$. We will show that $M$ is an eigenvalue of $B$. (The argument is similar when $m < 0$.)

First we prove that the supremum for $M$ is attained and that the corresponding $u$ is an eigenvector with eigenvalue $M$, meaning $Bu = Mu$.

Take a supremizing sequence $\{u_k\}$, normalized by $\|u_k\| = 1$, such that

$$\langle u_k, Bu_k\rangle \to M.$$

After passing to a subsequence we may suppose $u_k \rightharpoonup u$ weakly for some $u \in H$, by weak sequential compactness of the closed unit ball (Theorem 2.31). Then

$$\langle u_k, Bu_k\rangle \to \langle u, Bu\rangle$$

by Lemma 4.5, since $B$ is compact. Thus $M = \langle u, Bu\rangle$.

Notice $\|u\| \leq \liminf \|u_k\| = 1$ by Exercise 2.20, and conversely $\|u\| \geq 1$ because

$$M \geq \frac{\langle u, Bu\rangle}{\langle u, u\rangle} = \frac{M}{\langle u, u\rangle} > 0.$$

Hence $\|u\| = 1$, and so the supremum for $M$ is attained at $u$.

Now we show that this maximizing vector $u$ is an eigenvector with eigenvalue $M$. Fix $v \in H$ (the "variation direction") and define a real valued function

$$g(t) = \frac{\langle u + tv, B(u + tv)\rangle}{\langle u + tv, u + tv\rangle}$$

for small values of $t$ (so that the denominator is positive). We know $g$ is maximal at $t = 0$, by definition of $M$, and so

$$\begin{aligned} 0 &= g'(0) \\ &= \frac{d}{dt}\Big|_{t=0} \frac{\langle u, Bu\rangle + 2t\,\mathrm{Re}\langle v, Bu\rangle + t^2\langle v, Bv\rangle}{\langle u, u\rangle + 2t\,\mathrm{Re}\langle v, u\rangle + t^2\langle v, v\rangle} \\ &= \frac{d}{dt}\Big|_{t=0} \frac{M + 2t\,\mathrm{Re}\langle v, Bu\rangle + O(t^2)}{1 + 2t\,\mathrm{Re}\langle v, u\rangle + O(t^2)}. \end{aligned}$$

After evaluating the derivative we find

$$\mathrm{Re}\,\langle v, Bu\rangle - M\,\mathrm{Re}\,\langle v, u\rangle = 0.$$

For a real Hilbert space we can dispense with the "Re" part. We can dispense with to on a complex Hilbert space too, since the preceding formula holds also with $v$ replaced by $iv$, which changes the real part to an imaginary part. Hence we find

$$\langle v, Bu - Mu\rangle = 0$$

for all $v \in H$, so that $Bu = Mu$. Thus $u$ is an eigenvector with eigenvalue $M$, as we wanted to show.

Step 2 — Finding the largest-magnitude eigenvalue. Let

$$\alpha_1 = \sup_{u \neq 0} \frac{|\langle u, Bu\rangle|}{\langle u, u\rangle}$$

so that $\alpha_1$ equals the larger of $|m|$ and $|M|$. Step 1 shows that if $\alpha_1 \neq 0$ then B has a real eigenvalue $\beta_1 = \pm\alpha_1$. Denote the corresponding eigenvector by $u_1$, so that $Bu_1 = \beta_1 u_1$.

Step 3 — Repeat on the orthogonal complement. Let $H_1$ be the span of the eigenvector $u_1$ (that is, the subspace of all scalar multiples of $u_1$), so that $H_1$ is a closed subspace of $H$. We can decompose the Hilbert space as

$$H = H_1 \oplus H_1^{\perp}$$

by Theorem 2.16. Notice B maps $H_1^{\perp}$ to $H_1^{\perp}$, since if $w \in H_1^{\perp}$ then $\langle w, u_1\rangle = 0$ and so

$$\langle Bw, u_1\rangle = \langle w, Bu_1\rangle = \beta_1\langle w, u_1\rangle = 0,$$

where we used once again the selfadjointness of B.

Hence one sees that $H_1^{\perp}$ is a separable, infinite dimensional Hilbert space, and $B : H_1^{\perp} \to H_1^{\perp}$ is a linear, compact, selfadjoint operator. Thus we may apply Step

2 to the operator $B$ restricted to $H_1^\perp$, and in this fashion we continue iteratively generating eigenvalues $\beta_1, \beta_2, \beta_3, \dots$ with decreasing magnitudes $|\beta_1| \geq |\beta_2| \geq |\beta_3| \geq \cdots$. The corresponding eigenvectors are orthonormal by construction, and so

$$H = (H_1 \oplus H_2 \oplus \cdots \oplus H_k) \oplus (H_1 \oplus H_2 \oplus \cdots \oplus H_k)^\perp$$

for each $k$.

The process could terminate after finitely many iterations: precisely, it would terminate after $k$ iterations if $\alpha_{k+1} = 0$ where

$$\alpha_{k+1} = \sup_{u \perp (H_1 \oplus \cdots \oplus H_k)} \frac{|\langle u, Bu \rangle|}{\langle u, u \rangle}$$

Step 4 — Suppose the process terminates after $k$ iterations because $\alpha_{k+1} = 0$. Let $K = (H_1 \oplus \cdots \oplus H_k)^\perp$. Then by definition of $\alpha_{k+1}$ we have

$$\langle u, Bu \rangle = 0, \qquad u \in K. \tag{4.5}$$

Then

$$\langle v, Bu \rangle = 0, \qquad u, v \in K, \tag{4.6}$$

because

$$\begin{aligned} 2\,\mathrm{Re}\langle v, Bu \rangle &= \langle v, Bu \rangle + \langle Bu, v \rangle \\ &= \langle v, Bu \rangle + \langle u, Bv \rangle \qquad \text{by selfadjointness of } B \\ &= \langle u + v, B(u + v) \rangle - \langle u, Bu \rangle - \langle v, Bv \rangle \\ &= 0 \end{aligned}$$

by (4.5), and similarly for the imaginary part (after replacing $v$ with $iv$).

Choosing $v = Bu$ in (4.6) shows that $Bu = 0$ for all $u \in K$. Thus $K$ is the kernel or zero eigenspace of the operator $B$. Clearly $K$ has a countable ONB, since $H$ is separable by hypothesis. Combining this basis with the eigenvectors $u_1, \dots, u_k$ yields an ONB for the whole space $H$.

Step 5 — Suppose the process does not terminate. Then $\lim \beta_k = 0$, as follows. The orthonormal sequence $\{u_k\}$ converges weakly to $0$ by Exercise 2.21. Hence $Bu_k$ converges in norm to $0$, by Lemma 4.5, and so

$$|\beta_k| = \|\beta_k u_k\| = \|Bu_k\| \to 0.$$

Next we let $K = \left( \oplus_{k=1}^\infty H_k \right)^\perp$. If $K = \{0\}$ then the $\{u_k\}$ form an ONB for $H$, and the proof is complete.

Suppose $K \neq \{0\}$. Observe that $\langle u, Bu\rangle = 0$ for all $u \in K$ because

$$\frac{|\langle u, Bu\rangle|}{\langle u, u\rangle} \leq \alpha_{k+1} \to 0.$$

Hence $Bu = 0$ by arguing as in Step 4, and so we complete the proof by choosing an ONB for $K$ and combining it with the $\{u_k\}$. □

***Exercise*** 4.1 (Diagonal operator on $\ell^2$). Let $H = \ell^2(\mathbb{N})$ and suppose $\{\beta_k\}$ is a bounded sequence of real numbers. Write $x = (x_1, x_2, x_3, \dots) \in \ell^2(\mathbb{N})$ and define a "diagonal" operator $B : H \to H$ by

$$Bx = (\beta_1 x_1, \beta_2 x_2, \beta_3 x_3, \dots).$$

Obviously $\ell^2(\mathbb{N})$ has an ONB of eigenvectors of B, since $Be_k = \beta_k e_k$ where $e_k$ denotes the standard $k$-th unit vector

(i) Show B is a bounded linear operator, and that B is selfadjoint.

(ii) Show that if $\beta_k \to 0$ then B is compact.

***Exercise*** 4.2 (Hilbert–Schmidt integral operators are compact, sometimes self-adjoint). Let $U$ be a domain in $\mathbb{R}^N$ and consider a kernel $K \in L^2(U \times U)$. The functions in this exercise are real-valued.

(i) Show that the *Hilbert–Schmidt* operator $B : L^2(U) \to L^2(U)$ defined by

$$Bu(x) = \int_U K(x,y)u(y)\, dy, \qquad x \in U,$$

is bounded and linear, with operator norm $\|B\| \leq \|K\|_{L^2(U\times U)}$.

(ii) Since $L^2(U)$ is separable, it has a countable ONB, which we denote $\{v_j\}_{j\in\mathbb{N}}$. Define a kernel

$$K_n(x,y) = \sum_{j=1}^{n} \sum_{k=1}^{\infty} \kappa_{jk} v_j(x) v_k(y), \qquad x, y \in U, \quad n \geq 1,$$

where the coefficient is given by the inner product of the kernel with the ONB functions on the product space:

$$\kappa_{jk} = \int_U \int_U K(x,y) v_j(x) v_k(y)\, dx dy = \langle v_j, Bv_k\rangle_{L^2(U)}.$$

Consider the Hilbert–Schmidt operator $B_n$ having kernel $K_n$. Prove $B_n$ converges uniformly to B in the operator norm, meaning

$$\sup_u \frac{\|B_n u - Bu\|_{L^2(U)}}{\|u\|_{L^2(U)}} \to 0$$

as $n \to \infty$. *Hint.* Write $u = \sum_k c_k v_k$ and recall Parseval's identity.

(iii) Conclude from Exercise 1.8 that Hilbert–Schmidt operators are compact.

(iv) Show the adjoint $B^*$ is a Hilbert–Schmidt operator with kernel $K^*(x, y) = K(y, x)$. Hence if $K$ is symmetric then $B$ is selfadjoint.

(v) Assume $K$ is symmetric, so that $B$ generates an ONB of eigenfunctions, by the Spectral Theorem 4.6. Deduce that the spectral radius is bounded by the norm of the kernel, meaning $|\beta_k| \leq \|K\|_{L^2(U \times U)}$ for all $k$.

***Exercise*** 4.3 (Green's operator on the unit interval). Define *Green's operator* for the unit interval by

$$Gu(x) = \int_0^1 K(x, y)u(y)\, dy, \qquad x \in (0, 1),$$

where the kernel is

$$K(x, y) = \begin{cases} (1-x)y, & \text{when } 0 < y \leq x < 1, \\ x(1-y), & \text{when } 0 < x \leq y < 1. \end{cases}$$

Clearly Green's operator is an example of a Hilbert–Schmidt operator, as defined in Exercise 4.2. Hence by that exercise, $G : L^2(0, 1) \to L^2(0, 1)$ is bounded and linear, and is selfadjoint (since $K(x, y) = K(y, x)$). Therefore $L^2(0, 1)$ has an ONB consisting of eigenfunctions of $G$.

(i) Use the equation

$$Gu = \beta u$$

and the definition of $G$ to show that the eigenfunction $u$ is twice differentiable (classically) and satisfies $-u = \beta u''$. (As part of the proof, explain why $\beta \neq 0$.)

(ii) Note that $Gu(x)$ is continuous on $[0, 1]$. Show $Gu = 0$ at $x = 0, 1$, so that $u = 0$ at $x = 0, 1$. Hence $u$ satisfies Dirichlet boundary conditions.

(iii) [Necessary condition for eigenfunctions] Deduce $u(x) = c \sin(n\pi x)$ for some constant $c \neq 0$ and some $n \in \mathbb{N}$, in which case the eigenvalue is $\beta = 1/(n\pi)^2$.

(iv) [Sufficient condition for eigenfunctions] Show that $u(x) = \sin(n\pi x)$ is an eigenfunction of $G$, for each $n \in \mathbb{N}$.

(v) Conclude that $\{\sqrt{2} \sin(n\pi x)\}_{n \in \mathbb{N}}$ is an ONB for $L^2(0, 1)$.

## 4.3 *Application:* ONB of eigenfunctions for symmetric elliptic operator

Suppose $u(x)$ is defined on a bounded domain $U \subset \mathbb{R}^N$. Consider throughout this section the **divergence form symmetric elliptic operator**

$$\boxed{Lu = -\sum_{j,k=1}^{N} \left(a^{jk}(x)u_{x_j}\right)_{x_k} + c(x)u}$$

where the real-valued, measurable coefficient functions $a^{jk}$ are symmetric:

$$a^{jk} = a^{kj},$$

and **elliptic**:

$$\sum_{j,k=1}^{N} a^{jk}(x)\xi_j\xi_k \geq \theta|\xi|^2$$

for almost every $x \in U$ and all $\xi \in \mathbb{R}^N$, for some constant $\theta > 0$.

*Example* 4.7. The negative Laplacian arises from choosing $a^{jk} = 1$ when $j = k$ and $a^{jk} = 0$ when $j \neq k$, and $c \equiv 0$, because then $L = -\Delta$.

*Notes.* The operator $L$ is **symmetric**, or **formally selfadjoint**, because

$$\langle Lu, v\rangle_{L^2} = \langle u, Lv\rangle_{L^2}$$

by formal integration by parts, ignoring boundary terms. Obviously if $L$ included first order derivative terms (as we will consider in Chapter 5) then it would no longer be symmetric.

Ellipticity means that the real symmetric matrix $A = \left(a^{jk}(x)\right)_{j,k=1}^{N}$ is positive definite at almost every point of $U$, with smallest eigenvalue greater than or equal to $\theta$. In terms of this matrix, the operator can be written

$$\boxed{Lu = -\nabla \cdot \left(A\nabla u\right) + cu}.$$

In physical applications, the matrix $A$ encodes information about spatial and directional inhomogeneities of the medium in which diffusion or wave motion takes place.

We claim that symmetric elliptic operators generate ONBs of eigenfunctions. To state the result we use the sesquilinear form

$$\boxed{\alpha(u, v) = \int_U \Big(\sum_{j,k} a^{jk}u_{x_j}v_{x_k} + cuv\Big)\,dx}. \tag{4.7}$$

**Corollary 4.8** (ONB of Dirichlet eigenfunctions of L)**.** *Assume* $a^{jk}, c \in L^\infty(U)$. *Then functions* $u_1, u_2, u_3, \ldots \in H_0^1(U)$ *exist along with numbers*

$$(\text{essinf}\, c) < \lambda_1 \leq \lambda_2 \leq \lambda_3 \leq \cdots \to \infty$$

*such that:*

- $u_n$ *is a weak Dirichlet eigenfunction of* $L$ *with eigenvalue* $\lambda_n$

$$\begin{aligned} Lu_n &= \lambda_n u_n \quad \text{in } U, \\ u_n &= 0 \qquad \text{on } \partial U, \end{aligned}$$

  *in the weak sense that*

$$\alpha(u_n, v) = \lambda_n \langle u_n, v \rangle_{L^2}, \qquad v \in H_0^1(U),$$

  *or more explicitly that*

$$\int_U \big( \sum_{j,k} a^{jk} (u_n)_{x_j} v_{x_k} + c u_n v \big)\, dx = \lambda_n \int_U u_n v \, dx, \qquad v \in H_0^1(U);$$

- $\{u_n\}$ *is an ONB for* $L^2(U)$.

*The decomposition*

$$f = \sum_n \langle f, u_n \rangle_{L^2} u_n$$

*converges in* $L^2$ *for each* $f \in L^2(U)$, *and converges in* $H_0^1$ *for each* $f \in H_0^1(U)$.

*Proof.* We will apply the Discrete Spectral Theorem 4.2 with Hilbert spaces $\mathcal{H} = L^2(U)$ and $\mathcal{K} = H_0^1(U)$. Note $L^2(U)$ is separable. Clearly $H_0^1$ imbeds densely into $L^2$, because $C_c^\infty$ is dense in both of them. Further, $H_0^1$ imbeds compactly into $L^2$, by Corollary 3.28(i).

Define a sesquilinear form on $H_0^1(U)$ by

$$\begin{aligned} a(u,v) &= \alpha(u,v) + (\theta - \text{essinf}\, c)\langle u, v \rangle_{L^2} \\ &= \int_U \big( \sum_{j,k} a^{jk} u_{x_j} v_{x_k} + cuv + (\theta - \text{essinf}\, c) uv \big)\, dx. \end{aligned}$$

This form is bounded on $H_0^1 \times H_0^1$ since $a^{jk}$ and $c$ are bounded functions, and is symmetric since $a(u,v) = a(v,u)$ by symmetry of the $a^{jk}$. For ellipticity of $a(\cdot,\cdot)$, notice that

$$\begin{aligned} a(u,u) &= \int_U \big( \sum_{j,k} a^{jk} u_{x_j} u_{x_k} + (\theta + c - \text{essinf}\, c) u^2 \big)\, dx \\ &\geq \theta \int_U \big( |\nabla u|^2 + u^2 \big)\, dx = \theta \|u\|_{H^1}^2 \end{aligned}$$

by the ellipticity hypothesis on the $a^{jk}$.

Hence Theorem 4.2 yields functions $u_n$ and numbers $\gamma_n > 0$ such that

$$a(u_n, v) = \gamma_n \langle u_n, v \rangle_{L^2} \tag{4.8}$$

for all $v \in H_0^1(U)$. Define

$$\lambda_n = \gamma_n - (\theta - \operatorname{essinf} c). \tag{4.9}$$

Then the definition of $a(\cdot,\cdot)$ implies with the help of (4.8) that $\alpha(u_n, v) = \lambda_n \langle u_n, v \rangle_{L^2}$, which means $Lu_n = \lambda_n u_n$ weakly.

The remaining claims in the theorem follow from Theorem 4.2, except that we must still prove the lower bound $\lambda_1 > \operatorname{essinf} c$. For that, notice

$$\begin{aligned} \lambda_1 \langle u_1, u_1 \rangle_{L^2} &= \alpha(u_1, u_1) \qquad \text{since } Lu_1 = \lambda_1 u_1 \text{ weakly} \\ &\geq \theta \int_U |\nabla u_1|^2 \, dx + (\operatorname{essinf} c) \langle u_1, u_1 \rangle_{L^2} \end{aligned}$$

by ellipticity. The first term on the right is positive, since it is bounded below by a constant times $\int_U u_1^2 \, dx$, by Exercise 3.16. Hence $\lambda_1 > \operatorname{essinf} c$. □

Next we consider the Neumann spectrum. Write $n$ for the outward normal vector. To avoid confusion, we will label our eigenfunctions with "$m$" in what follows, rather than with "$n$". Note also that the definition of weak Neumann eigenfunction, in (4.10) below, uses trial functions in all of $H^1$, not just in $H_0^1$ as used for the Dirichlet eigenfunctions.

**Corollary 4.9** (ONB of Neumann eigenfunctions of $L$)**.** *Assume $a^{jk}, c \in L^\infty(U)$, and that $\partial U$ is $C^1$-smooth. Then functions $v_1, v_2, v_3, \ldots \in H^1(U)$ exist along with numbers*

$$(\operatorname{essinf} c) \leq \mu_1 \leq \mu_2 \leq \mu_3 \leq \cdots \to \infty$$

*such that:*

- *$v_m$ is a weak Neumann eigenfunction of $L$ with eigenvalue $\mu_m$*

  $$\begin{aligned} Lv_m &= \mu_m v_m \quad \text{in } U, \\ (A\nabla v_m) \cdot n &= 0 \qquad \text{on } \partial U, \end{aligned}$$

  *in the weak sense that*

  $$\alpha(v_m, w) = \mu_m \langle v_m, w \rangle_{L^2}, \qquad w \in H^1(U),$$

  *or more explicitly that*

  $$\int_U \Big( \sum_{j,k} a^{jk} (v_m)_{x_j} w_{x_k} + c v_m w \Big) \, dx = \mu_m \int_U v_m w \, dx, \qquad w \in H^1(U); \tag{4.10}$$

- $\{v_m\}$ *is an ONB for* $L^2(U)$.

*The decomposition*

$$f = \sum_m \langle f, v_m \rangle_{L^2} v_m$$

*converges in* $L^2$ *for each* $f \in L^2(U)$, *and converges in* $H^1$ *for each* $f \in H^1(U)$.

***Exercise*** 4.4. Prove Corollary 4.9.

***Exercise*** 4.5 (Neumann boundary conditions arise "naturally" from the weak eigenfunction equation). Show that if the Neumann eigenfunction $v_m$ and the coefficients $a^{jk}$ are smooth on $\overline{U}$, then

$$(A\nabla v_m) \cdot n = 0 \qquad \text{on } \partial U.$$

That is, the eigenfunction satisfies the Neumann boundary condition in the classical sense. *Hint.* Use the definition (4.10) of weak solution, and the fact that $Lv_m = \mu_m v_m$ classically, since $v_m$ is assumed to be smooth.

*Example* 4.10. In one dimension on the interval $U = (0, \pi)$, the second derivative operator $L = -\frac{d^2}{dx^2}$ has Dirichlet eigenfunctions and eigenvalues

$$u_n(x) = \sqrt{\frac{2}{\pi}} \sin(nx), \qquad \lambda_n = n^2, \qquad n = 1, 2, 3, \ldots,$$

and has Neumann eigenfunctions and eigenvalues

$$v_1(x) = \sqrt{\frac{1}{\pi}}, \qquad \mu_1 = 0,$$

and

$$v_m(x) = \sqrt{\frac{2}{\pi}} \cos\big((m-1)x\big), \qquad \mu_m = (m-1)^2, \qquad m = 2, 3, \ldots.$$

To justify these claims, one solves the weak eigenfunction equation $-u'' = \lambda u$ along with the Dirichlet boundary condition $u(0) = 0 = u(\pi)$, noting that weak eigenfunctions are classical eigenfunctions (by regularity theory as in Exercise 5.4). The functions $\sin(nx)$ for $n \geq 1$ are the only solutions. For the Neumann eigenfunctions one argues similarly, using the boundary condition $v'(0) = 0 = v'(\pi)$.

The spectral theory in this chapter applies also to higher order operators such as the biLaplacian $\Delta^2 = \Delta\Delta$, which arises in the wave equation for a vibrating rod (1 dimension) or vibrating plate (2 dimensions).

**Corollary 4.11** (ONB of Dirichlet eigenfunctions of biLaplacian)**.** *Functions* $u_1, u_2, u_3, \ldots \in H_0^2(U)$ *exist along with numbers*

$$0 < \Gamma_1 \leq \Gamma_2 \leq \Gamma_3 \leq \cdots \to \infty$$

*such that:*

- $u_n$ *is a weak Dirichlet eigenfunction of* $\Delta^2$ *with eigenvalue* $\Gamma_n$*:*

$$\begin{aligned} \Delta^2 u_n &= \Gamma_n u_n \quad \text{in } U, \\ u_n = |\nabla u_n| &= 0 \qquad \text{on } \partial U, \end{aligned}$$

  *in the weak sense that:*

$$\int_U \Delta u_n \Delta v \, dx = \Gamma_n \int_U u_n v \, dx, \qquad v \in H_0^2(U);$$

- $\{u_n\}$ *is an ONB for* $L^2(U)$.

*The decomposition*

$$f = \sum_n \langle f, u_n \rangle_{L^2} u_n$$

*converges in* $L^2$ *for each* $f \in L^2(U)$*, and converges in* $H_0^2$ *for each* $f \in H_0^2(U)$.

***Exercise*** 4.6*.* Prove Corollary 4.11.

## 4.4 Lax–Milgram and nonsymmetric sesquilinear forms

In the next chapter we want to state results on *non-selfadjoint* differential operators, such as operators with first order derivative terms. For these operators the Lax–Milgram Theorem plays the role of the Riesz Representation Theorem, and so we state the theorem below along with an existence and uniqueness corollary. This section can be omitted, since we will not explicitly use it in later proofs.

**Theorem 4.12** (Lax–Milgram)**.** *Consider a complex Hilbert space* $\mathcal{K}$ *and a sesquilinear form* $a : \mathcal{K} \times \mathcal{K} \to \mathbb{C}$ *that is continuous and elliptic (but not necessarily symmetric, so that* $a(v, u) \neq \overline{a(u, v)}$ *in general).*

*Then a bounded linear bijection* $A : \mathcal{K} \to \mathcal{K}$ *exists with* $A^{-1}$ *bounded and*

$$a(v, u) = \langle v, Au \rangle_{\mathcal{K}}, \qquad u, v \in \mathcal{K}. \tag{4.11}$$

The point of the theorem is to express the sesquilinear form in terms of the inner product. *Note.* The theorem holds also for real Hilbert spaces and sesquilinear forms, simply without the complex conjugates.

*Proof of Theorem 4.12.* For each fixed $u \in \mathcal{K}$, the map $v \mapsto a(v, u)$ is a bounded linear functional on $\mathcal{K}$ and so by the Riesz Representation Theorem 2.19, a unique vector $Au \in K$ exists such that (4.11) holds for all $v \in \mathcal{K}$.

Linearity of the map $u \mapsto Au$ is easily checked by the definition. To show $A$ is bounded, note that

$$\|Au\|_{\mathcal{K}}^2 = \langle Au, Au\rangle_{\mathcal{K}} = a(Au, u) \leq (\text{const.})\|Au\|_{\mathcal{K}}\|u\|_{\mathcal{K}}$$

by boundedness of the sesquilinear form, and so $\|Au\|_{\mathcal{K}} \leq (\text{const.})\|u\|_{\mathcal{K}}$.

To obtain a bound in the opposite direction, observe that

$$\begin{aligned} c\|u\|_{\mathcal{K}}^2 &\leq a(u, u) && \text{by ellipticity} \\ &= \langle u, Au\rangle_{\mathcal{K}} && \text{by choosing } v = u \text{ in (4.11)} \\ &\leq \|u\|_{\mathcal{K}}\|Au\|_{\mathcal{K}} && \text{by Cauchy–Schwarz,} \end{aligned} \tag{4.12}$$

so that

$$c\|u\|_{\mathcal{K}} \leq \|Au\|_{\mathcal{K}}. \tag{4.13}$$

Thus if $Au = 0$ then $u = 0$, and so $A$ is injective.

Next we show $A$ is surjective. To start with, the range of $A$ is closed because if $Au_k \to w \in K$ then $\{Au_k\}$ is a Cauchy sequence and so $\{u_k\}$ is Cauchy by (4.13), and hence $u_k \to u$ for some $u \in \mathcal{K}$, so that $w = \lim Au_k = Au \in R(A) = \text{range}\,(A)$. Then since the range is a closed subspace we can orthogonally decompose the space as $K = R(A) \oplus R(A)^\perp$. In fact, $R(A)^\perp$ consists of only the zero vector, since if $z \in R(A)^\perp$ then by estimate (4.12),

$$c\|z\|_{\mathcal{K}}^2 \leq \langle z, Az\rangle_{\mathcal{K}} = 0$$

and so $z = 0$. Hence $R(A)^\perp = \{0\}$ and so $\mathcal{K} = R(A)$, meaning $A$ is surjective.

Lastly, since $A : \mathcal{K} \to \mathcal{K}$ is a linear bijection, we know $A^{-1} : \mathcal{K} \to \mathcal{K}$ is also linear, and it is bounded due to (4.13):

$$\|A^{-1}u\|_{\mathcal{K}} \leq \frac{1}{c}\|u\|_{\mathcal{K}}.$$

□

Now we develop an existence and uniqueness result for sesquilinear forms that need not be symmetric, thereby generalizing Step 1 in the proof of Theorem 4.2.

**Corollary 4.13** (Solution map for sesquilinear forms). *Assume $\mathcal{K}$ is continuously imbedded in a complex Hilbert space $\mathcal{H}$. Then for each $f \in \mathcal{H}$, a unique $u \in \mathcal{K}$ exists such that*

$$a(v, u) = \langle v, f \rangle_{\mathcal{H}}, \qquad v \in \mathcal{K}.$$

*Further, the* solution map

$$\begin{aligned} B : &\mathcal{H} \to \mathcal{K} \\ &f \mapsto u \end{aligned}$$

*is linear and bounded.*

*If in addition $\mathcal{K}$ imbeds compactly into $\mathcal{H}$, then the solution map $B : \mathcal{H} \to \mathcal{H}$ is compact.*

*Proof of Corollary 4.13.* Let $f \in \mathcal{H}$. The functional $v \mapsto \langle v, f \rangle_{\mathcal{H}}$ is bounded and linear on $\mathcal{K}$, since for all $v \in \mathcal{K}$ we have that

$$\begin{aligned} \left| \langle v, f \rangle_{\mathcal{H}} \right| &\leq \|v\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \\ &\leq (\text{const.}) \|v\|_{\mathcal{K}} \|f\|_{\mathcal{H}} \end{aligned}$$

using that $\mathcal{K}$ imbeds into $\mathcal{H}$. Hence by Riesz Representation, a unique vector $F \in \mathcal{K}$ exists such that

$$\langle v, f \rangle_{\mathcal{H}} = \langle v, F \rangle_{\mathcal{K}}, \qquad v \in \mathcal{K}.$$

Letting $u = A^{-1}F$, we find

$$\begin{aligned} a(v, u) &= \langle v, Au \rangle_{\mathcal{K}} \qquad \text{by the Lax–Milgram Theorem 4.12} \\ &= \langle v, F \rangle_{\mathcal{K}} \\ &= \langle v, f \rangle_{\mathcal{H}} \end{aligned}$$

as desired. Further, the map $B : f \mapsto F \mapsto A^{-1}F$ is bounded and linear from $\mathcal{H}$ to $\mathcal{K}$. Uniqueness of $u$ is easily proved using Lax–Milgram, and the final statement of the corollary is immediate. □

# Chapter 5

# Second order linear elliptic PDEs: existence and regularity of solutions

### Introduction

One of the key insights in in PDE theory in the last century was that existence of solutions should be decoupled from the question of their regularity. Thus in this chapter we first prove a solution exists weakly, and then we show that under suitable hypotheses on the coefficient functions, the weak solution is smooth and hence classical.

Take $U$ to be a bounded domain in $\mathbb{R}^N$.

## 5.1 Generalized Poisson equation

Throughout the section, $L$ is a second order linear elliptic operator in divergence form:

$$\boxed{Lu = -\sum_{j,k=1}^{N} \left(a^{jk}(x)u_{x_j}\right)_{x_k} + \sum_j b^j(x)u_{x_j} + c(x)u},$$

where we always assume that $a^{jk} = a^{kj}$ and the $a^{jk}$ are elliptic with constant $\theta > 0$ that is independent of $x \in U$. Section 4.3 treated the special case of symmetric operators ($b^j \equiv 0$), and we constructed ONBs of eigenfunctions for such formally selfadjoint operators. Note one has no reason to expect such eigenfunctions to exist for the general non-symmetric $L$.

Given $f \in L^2(U)$, we say $u \in H^1_0(U)$ is a **weak solution** of the generalized Poisson equation with Dirichlet boundary condition

$$\begin{aligned} Lu &= f \quad \text{in } U, \\ u &= 0 \quad \text{on } \partial U, \end{aligned}$$

if

$$\alpha(u, v) = \langle f, v \rangle_{L^2}, \qquad v \in H^1_0(U),$$

where the sesquilinear form is

$$\boxed{\alpha(u, v) = \int_U \Big( \sum_{j,k} a^{jk} u_{x_j} v_{x_k} + \sum_j b^j u_{x_j} v + cuv \Big)\, dx}.$$

Notice $\alpha(u, v) \neq \alpha(v, u)$ in general, due to lack of symmetry of the term $b^j u_{x_j} v$.

We say $L$ has a **zero eigenvalue** if the equation $Lu = 0$ (meaning, with $f \equiv 0$) has a weak solution $u \not\equiv 0$; that is, if $\alpha(u, v) = 0$ for all $v \in H^1_0(U)$.

The next theorem shows wellposedness of this boundary value problem, assuming $L$ does not have a zero eigenvalue.

**Theorem 5.1** (Existence, uniqueness, and continuous dependence on the data). *Assume $a^{jk}, b^j, c \in L^\infty(U)$ and suppose $L$ does not have a zero eigenvalue.*

*If $f \in L^2(U)$ then a unique $u \in H^1_0(U)$ exists such that*

$$\begin{aligned} Lu &= f \quad \textit{in } U, \\ u &= 0 \quad \textit{on } \partial U, \end{aligned}$$

*weakly. Further, the solution map $L^{-1} : L^2(U) \to H^1_0(U)$ defined by $L^{-1} f = u$ is linear and continuous (bounded), with*

$$\|u\|_{H^1(U)} \leq C \|f\|_{L^2(U)}$$

*where $C = C(U, L)$. Hence also $L^{-1} : L^2(U) \to L^2(U)$ is compact.*

*Proof for symmetric $L$, with $b^j \equiv 0$.*
(For the general case see [Evans], where the treatment is based on the Lax–Milgram Theorem 4.12 instead of an ONB of eigenfunctions as we use below.)

Step 1 — Uniqueness. Assume $u$ is a weak solution. For each eigenfunction $u_n$ (constructed in Corollary 4.8) we have

$$\begin{aligned} \langle u, u_n \rangle_{L^2} &= \frac{1}{\lambda_n} \alpha(u, u_n) \qquad \text{since } Lu_n = \lambda_n u_n \text{ weakly and } \lambda_n \neq 0 \text{ by hypothesis,} \\ &= \frac{1}{\lambda_n} \langle f, u_n \rangle_{L^2} \end{aligned}$$

since $Lu = f$ weakly. Hence the orthonormal decomposition $u = \sum_n \langle u, u_n \rangle_{L^2} u_n$ implies

$$u = \sum_n \frac{1}{\lambda_n} \langle f, u_n \rangle_{L^2} u_n. \tag{5.1}$$

This formula determines $u$ uniquely in terms of $f$.

Step 2 — Existence. Define $u$ by (5.1); we will show below that this series converges in $H^1_0$. Then

$$a(u, v) = \sum_n \frac{1}{\lambda_n} \langle f, u_n \rangle_{L^2} \, a(u_n, v)$$

since $a(\cdot, v)$ is continuous on $H^1$ and the series (5.1) converges in $H^1$. Thus

$$\begin{aligned} a(u, v) &= \sum_n \langle f, u_n \rangle_{L^2} \, \langle u_n, v \rangle_{L^2} \qquad \text{since } Lu_n = \lambda_n u_n \text{ weakly} \\ &= \langle f, v \rangle_{L^2} \end{aligned}$$

since $\sum_n \langle f, u_n \rangle_{L^2} u_n = f$ in $L^2$. Hence $Lu = f$ weakly.

To prove that the series (5.1) defining $u$ converges in $H^1_0$, we first rewrite it as

$$u = \sum_n \frac{\sqrt{\gamma_n}}{\lambda_n} \langle f, u_n \rangle_{L^2} \frac{u_n}{\sqrt{\gamma_n}} \tag{5.2}$$

where $\gamma_n$ and $\lambda_n$ are related by (4.9). We know from the application of the Discrete Spectral Theorem 4.2 in the last chapter that $\{u/\sqrt{\gamma_n}\}$ forms an ONB for $H^1_0$, with respect to the $a$-inner product. Hence we need only show the coefficient sequence in (5.2) belongs to $\ell^2$, because then the Synthesis Proposition 2.24 yields convergence of the series in $H^1_0$.

The numbers $\gamma_n$ and $\lambda_n$ tend to infinity, and differ only by a constant. In particular, $\sqrt{\gamma_n}/\lambda_n$ is less than 1 for all large $n$. Also the sequence $\{\langle f, u_n \rangle_{L^2}\}_{n=1}^\infty$ belongs to $\ell^2$ by Bessel's inequality, since the $u_n$ are $L^2$-orthonormal. Thus the coefficient sequence in (5.2) belongs to $\ell^2$.

Step 3 — Continuous dependence on the data. The definition (5.1) reveals that $u$ depends linearly on $f$. Further, its $L^2$-norm is controlled by the $L^2$-norm of $f$, with

$$\begin{aligned} \|u\|^2_{L^2} &= \sum_n \frac{1}{\lambda_n^2} \big| \langle f, u_n \rangle_{L^2} \big|^2 \qquad \text{by Parseval,} \\ &\leq \frac{1}{(\min_j |\lambda_j|)^2} \sum_n |\langle f, u_n \rangle_{L^2}|^2 \qquad (5.3) \\ &= C \|f\|^2_{L^2}. \end{aligned}$$

Thus $L^{-1}: L^2(U) \to L^2(U)$ is bounded.

Further, the $H^1$-norm of $u$ is controlled by the $L^2$-norm of $f$, with

$$\begin{aligned}
\|u\|_{H^1}^2 &\leq Ca(u,u) && \text{by ellipticity of the sesquilinear form } a \\
&\leq C\big(\alpha(u,u) + \|u\|_{L^2}^2\big) && \text{by definition of } a \text{ and } \alpha \\
&= C\big(\langle f, u\rangle_{L^2} + \|u\|_{L^2}^2\big) && \text{since } Lu = f \text{ weakly} \\
&\leq C\big(\|f\|_{L^2}\|u\|_{L^2} + \|u\|_{L^2}^2\big) && \text{by Cauchy–Schwarz} \\
&\leq C\|f\|_{L^2}^2
\end{aligned}$$

by the $L^2$-bound above. Thus $L^{-1}: L^2(U) \to H_0^1(U)$ is bounded.

Lastly, $H_0^1$ imbeds compactly into $L^2$ by Rellich–Kondrachev Corollary 3.28(i), and so the solution map $L^{-1}: L^2(U) \to L^2(U)$ is compact. □

*Remark* 5.2 (Blow-up). The result must fail if one of the eigenvalues equals $0$: if $\lambda_n = 0$ then $Lu_n = \lambda_n u_n = 0$, and clearly the norm of the solution $u_n$ cannot be bounded by the norm of the data $0$. Thus the bound (5.3) on the solution operator must blow up if the operator is perturbed to drive one of the eigenvalues to $0$ (say, by subtracting a suitable constant from the coefficient function $c(x)$ in $L$).

Now we treat the case of a zero eigenvalue. We need the **formal adjoint** of $L$, which is

$$L^*u = -\sum_{j,k} \left(a^{jk} u_{x_k}\right)_{x_j} - \sum_j \left(b^j u\right)_{x_j} + cu.$$

In the symmetric case ($b^j \equiv 0$) one has $L^* = L$. In the general case, the adjoint relation $\langle Lu, v\rangle_{L^2} = \langle u, L^*v\rangle_{L^2}$ holds after one integrates by parts formally and ignores boundary terms.

Given $f \in L^2(U)$, we define $u \in H_0^1(U)$ to be a weak solution of

$$\begin{aligned} L^*u &= f \quad \text{in } U, \\ u &= 0 \quad \text{on } \partial U, \end{aligned}$$

if $\alpha(v,u) = \langle v, f\rangle_{L^2}$ for all $v \in H_0^1(U)$. Then we say $L^*$ has a **zero eigenvalue** if the equation $L^*u = 0$ has a weak solution $u \not\equiv 0$, meaning $\alpha(v,u) = 0$ for all $v \in H_0^1(U)$.

**Theorem 5.3** (Existence and continuous dependence)**.** *Assume* $a^{jk}, b^j, c \in L^\infty(U)$ *and suppose* $L$ *has a zero eigenvalue. Then the homogeneous boundary value problem*

$$\begin{aligned} Lu &= 0 \quad \textit{in } U, \\ u &= 0 \quad \textit{on } \partial U, \end{aligned}$$

*has a nontrivial solution.*

*Further, for each* $f \in L^2(U)$ *the nonhomogeneous boundary value problem*

$$\begin{aligned} Lu &= f \quad \textit{in } U, \\ u &= 0 \quad \textit{on } \partial U, \end{aligned}$$

*has a weak solution* $u \in H_0^1(U)$ *if and only if* $\langle f, v \rangle_{L^2} = 0$ *for all* $0$*-eigenfunctions* $v$ *of* $L^*$*, that is, for each* $v \in H_0^1(U)$ *such that* $L^* v = 0$ *weakly. In that case*

$$\|u\|_{H^1(U)} \leq C \|f\|_{L^2(U)}$$

*where* $C = C(U, L)$*.*

*Proof for symmetric* $L$*, with* $b^j \equiv 0$*, so that* $L^* = L$*. For the general case see [Evans].*
Step 1 $\Longrightarrow$. Assume $Lu = f$ weakly. If $\lambda_n = 0$ is one of the zero eigenvalues of $L^* = L$, then

$$\langle f, u_n \rangle_{L^2} = a(u, u_n) = \lambda_n \langle u, u_n \rangle_{L^2} = 0$$

since $Lu_n = \lambda_n u_n$ weakly and $\lambda_n = 0$

Step 2 $\Longleftarrow$. Assume $\langle f, u_n \rangle_{L^2} = 0$ whenever $\lambda_n = 0$. Argue as in Steps 2 and 3 of the proof of Theorem 5.1 to show the existence of a solution $u$ with norm bounded by the $L^2$-norm of $f$; simply omit from the series all terms for which $\lambda_n = 0$. □

*Remark* 5.4 (Matrix analogy for Theorems 5.1 and 5.3). Let $M$ be a symmetric real matrix. If $M$ has no zero eigenvalue then $M$ is invertible and the vector equation $Mx = b$ has a unique solution, for each $b$. The solution can be written $x = M^{-1} b = \sum \mu_n^{-1} (b \cdot v_n) v_n$ where $\mu_n$ is the eigenvalue and $v_n$ is the eigenvector.

For a general matrix, not necessarily symmetric, the equation $Mx = b$ is solvable if and only if $b$ lies in the column space of $M$, which is the row space of the transpose $M^T$, and hence is the orthogonal complement of the kernel of $M^T$. That kernel is precisely the $0$-eigenspace of $M^T$.

The matrix case and our theorems on solvability of elliptic PDEs can be viewed as applications of the Fredholm Alternative for compact perturbations of the identity; see [Evans].

***Exercise*** 5.1 (Existence, uniqueness and continuous dependence under Neumann boundary conditions). Assume $a^{jk} \in L^\infty(U)$ and $b^j \equiv 0, c \equiv 0$, and suppose $\partial U$ is $C^1$-smooth.

We say $u \in H^1(U)$ is a weak solution for the Neumann boundary value problem

$$\begin{aligned} Lu &= f \quad \text{in } U, \\ (A \nabla u) \cdot n &= 0 \quad \text{on } \partial U, \end{aligned} \tag{5.4}$$

if

$$\int_U \sum_{j,k} a^{jk} u_{x_j} v_{x_k} \, dx = \int_U f v \, dx, \qquad v \in H^1(U).$$

(Notice the trial function $v$ ranges over all of $H^1$, not just $H^1_0$.)

(i) Show that the Neumann eigenvalues of $L$ satisfy $0 = \mu_1 < \mu_2$.
(ii) Prove that for $f \in L^2(U)$, the nonhomogeneous Neumann boundary value problem (5.4) has a weak solution $u \in H^1(U)$ if and only if

$$\int_U f \, dx = 0.$$

(iii) Show when $\int_U f \, dx = 0$ that the solution is unique up to additive constants.
(iv) Normalize the solution by requiring $\int_U u \, dx = 0$. Show that under this normalization, the solution map $L^{-1} : L^2(U) \to H^1(U)$ defined by $L^{-1} f = u$ is linear and bounded, with

$$\|u\|_{H^1(U)} \le C \|f\|_{L^2(U)}$$

where $C = C(U, L)$. Show also $L^{-1} : L^2(U) \to L^2(U)$ is compact.

***Exercise*** 5.2 (Existence, uniqueness and continuous dependence for the biLaplacian). We say $u \in H^2_0(U)$ is a weak solution for the biLaplacian boundary value problem

$$\begin{aligned} \Delta^2 u = f \quad & \text{in } U, \\ u = |\nabla u| = 0 \quad & \text{on } \partial U, \end{aligned} \tag{5.5}$$

if

$$\int_U \Delta u \Delta v \, dx = \int_U f v \, dx, \qquad v \in H^2_0(U).$$

(i) Prove that if $f \in L^2(U)$ then the biLaplacian boundary value problem (5.5) has a unique weak solution $u \in H^2_0(U)$.
(ii) Show that the solution map $\Delta^{-2} : L^2(U) \to H^2_0(U)$ defined by $\Delta^{-2} f = u$ is linear and bounded, with

$$\|u\|_{H^2(U)} \le C \|f\|_{L^2(U)}$$

where $C = C(U)$. Hence also $\Delta^{-2} : L^2(U) \to L^2(U)$ is compact.

*Remark* 5.5. A nonzero boundary condition can be handled by subtraction, as follows. Assume $h(x)$ is defined for all $x \in \overline{U}$ (not just for $x \in \partial U$). To solve the elliptic PDE $Lu = f$ with nonzero boundary condition $u = h$ on $\partial U$, we let

$\widetilde{u} = u - h$ and solve the corresponding elliptic problem for $\widetilde{u}$, which has zero boundary condition:

$$\begin{aligned} L\widetilde{u} &= \widetilde{f} \stackrel{\text{def}}{=} f - Lh \quad \text{in } U, \\ \widetilde{u} &= 0 \qquad\qquad\quad \text{on } \partial U. \end{aligned}$$

Of course, we need suitable hypotheses on $h$ so that this new problem fits our assumptions above.

## 5.2 Regularity of solutions

The modern theory of PDEs splits the question "Does a smooth solution exist?" into two parts: "Does a weak solution exist?" and "Is the weak solution smooth?" We have answered the first question already, for the divergence form elliptic equation $Lu = f$, and in this section we answer the second question.

The point is that if the weak solution $u$ is smooth, then it solves $Lu = f$ classically, which means we have answered the original question.

### Interior regularity

#### Motivation: an $H^2$-estimate for Poisson's equation

Assume $u \in C_c^\infty(\mathbb{R}^N)$ is a classical solution of Poisson's equation $-\Delta u = f$. The PDE shows that the data $f$ controls the sum of the pure second partial derivatives of $u$ (namely $\Delta u$). Less obviously, $f$ also controls the mixed derivatives of $u$:

$$\begin{aligned} \int f^2 \, dx &= \int (\Delta u)^2 \, dx \\ &= \sum_{j,k} \int u_{x_j x_j} u_{x_k x_k} \, dx \\ &= \sum_{j,k} \int u_{x_j x_k} u_{x_j x_k} \, dx \qquad \text{after integrating by parts twice} \\ &= \int \sum_{j,k} u_{x_j x_k}^2 \, dx \\ &= \int |D^2 u|^2 \, dx. \end{aligned}$$

Using arguments like above, we will show that weak solutions belong not only to $H^1$, but in fact to $H^2$. For this section we continue to assume $L$ is in divergence form:

$$\boxed{Lu = -\sum_{j,k} \left(a^{jk} u_{x_j}\right)_{x_k} + \sum_j b^j u_{x_j} + cu}$$

where the $a^{jk}$ are uniformly elliptic and symmetric ($a^{jk} = a^{kj}$).

We start by proving $H^2$-regularity away from the boundary. Our treatment follows [Evans, Chapter 6].

**Theorem 5.6** (Interior $H^2$-regularity). *Assume* $a^{jk} \in C^1 \cap L^\infty(U)$ *and* $b^j, c \in L^\infty(U)$*, and* $f \in L^2(U)$*.*

*If* $u \in H^1(U)$ *solves* $Lu = f$ *weakly (meaning* $\alpha(u, v) = \langle f, v \rangle_{L^2}$ *for all* $v \in H_0^1(U)$*) then* $u \in H^2_{loc}(U)$*. Further, on each open set* $V \Subset U$ *we have an* $H^2$*-norm estimate*

$$\|u\|_{H^2(V)} \leq C(\|f\|_{L^2(U)} + \|u\|_{L^2(U)})$$

*where* $C = C(U, V, L)$*.*

To explain the need for the $L^2$-norm of $u$ appearing on the right side of the estimate, notice we do not impose any boundary condition on $u$ — and so adding an arbitrarily large constant to $u$ might still yield a solution, which can cause the left side of the estimate to blow up; this reasoning holds for Poisson's equation, for example.

*Proof for special case* $L = -\Delta$ *on* $\mathbb{R}^N$*. (This method extends to the general case [Evans].)*
Step 1 — Basic properties of the difference operator. Fix the index $l = 1, \ldots, N$, and consider the first difference operator in the $l$-th direction:

$$D_h^l g(x) = \frac{g(x + he_l) - g(x)}{h}.$$

This difference operator commutes with derivatives, and so in particular

$$\nabla D_h^l = D_h^l \nabla.$$

Also, the difference operator satisfies an analogue of integration by parts, with

$$\int g(D_{-h}^l \widetilde{g})\, dx = -\int (D_h^l g)\widetilde{g}\, dx \tag{5.6}$$

whenever the functions $g, \widetilde{g}$ belong to $L^2$ (or to $L^p$ and $L^{p'}$ respectively), because:

$$\begin{aligned}\int g(D^l_{-h}\widetilde{g})\,dx &= -\frac{1}{h}\int g(x)\widetilde{g}(x-he_l)\,dx + \frac{1}{h}\int g(x)\widetilde{g}(x)\,dx\\ &= -\frac{1}{h}\int g(x+he_l)\widetilde{g}(x)\,dx + \frac{1}{h}\int g(x)\widetilde{g}(x)\,dx\\ &\qquad\qquad \text{by } x \mapsto x+he_l \text{ in the first integral}\\ &= -\int (D^l_h g)\widetilde{g}\,dx.\end{aligned}$$

Step 2 — Bounding the differences of the gradient. Assume $u \in H^1_0(\mathbb{R}^N)$ satisfies $Lu = f$ weakly. Choose $v = D^l_{-h}D^l_h u \in H^1_0(\mathbb{R}^N)$, with this choice motivated by the observation that $v$ "behaves like" $u_{x_l x_l}$, when $h$ is small. Then

$$\begin{aligned}\alpha(u,v) &= \langle f, v\rangle_{L^2}\\ \int \nabla u \cdot D^l_{-h}(D^l_h \nabla u)\,dx &= \int f D^l_{-h}D^l_h u\,dx\\ \int |D^l_h \nabla u|^2\,dx &= -\int f D^l_{-h}D^l_h u\,dx\end{aligned}$$

by "integration by parts"(5.6) on the left side. Estimating the right side with Cauchy–Schwarz gives that

$$\begin{aligned}\int |D^l_h \nabla u|^2\,dx &\le \|f\|_{L^2}\|D^l_{-h}(D^l_h u)\|_{L^2}\\ &\le \|f\|_{L^2}\|(D^l_h u)_{x_l}\|_{L^2}\end{aligned}$$

by Proposition 5.7(i) below. Thus

$$\|D^l_h \nabla u\|^2_{L^2} \le \|f\|_{L^2}\|D^l_h \nabla u\|_{L^2}$$

and simplifying gives

$$\|D^l_h \nabla u\|_{L^2} \le \|f\|_{L^2}.$$

Step 3 — Taking the limit of difference quotients. The previous formula implies that for each $k$, the difference quotients of $u_{x_k}$ are bounded in $L^2$, independently of the step size $h$. Hence $u_{x_k}$ is weakly differentiable by Proposition 5.7(ii) below, with

$$\|u_{x_k x_l}\|_{L^2} \le \|f\|_{L^2}.$$

□

The proof above relies on the following estimates between difference quotients and derivatives. Fix the index $l = 1, \ldots, N$, and recall the first difference operator in the $l$-th direction:

$$D_h^l g(x) = \frac{g(x + he_l) - g(x)}{h}.$$

**Proposition 5.7** (Difference quotients and derivatives)**.**
*(i) [Difference quotients are bounded by derivatives] Let* $1 \leq p < \infty$ *and* $h \neq 0$. *Then*

$$\|D_h^l u\|_{L^p} \leq \|u_{x_l}\|_{L^p}, \qquad u \in W_0^{1,p}(\mathbb{R}^N).$$

*(ii) [Difference quotient bounds imply derivative bounds] Let* $1 < p < \infty$ *and suppose* $u \in L^p(\mathbb{R}^N)$ *has compact support. Suppose* $\|D_h^l u\|_{L^p} \leq K$ *for all* $h \neq 0$. *Then* $u_{x_l}$ *exists weakly, and* $\|u_{x_l}\|_{L^p} \leq K$.

*Proof.*
(i) By density we may assume $u \in C_c^\infty(\mathbb{R}^N)$. The Fundamental Theorem implies that

$$\begin{aligned} |u(x + he_l) - u(x)| &= \Big| \int_0^1 \frac{d}{dt} u(x + the_l)\, dt \Big| \\ &\leq |h| \int_0^1 |u_{x_l}(x + the_l)|\, dt \\ &\leq |h| \Big( \int_0^1 |u_{x_l}(x + the_l)|^p\, dt \Big)^{1/p} \end{aligned}$$

by Hölder. Taking the $p$-th power and integrating with respect to $x$ yields

$$\begin{aligned} \int |D_h^l u|^p\, dx &\leq \int_0^1 \int |u_{x_l}(x + the_l)|^p\, dx dt \\ &= \int |u_{x_l}(x)|^p\, dx \end{aligned}$$

by changing variable with $x \mapsto x - the_l$.
(ii) Boundedness of the family $\{D_h^l u : h \neq 0\}$ in $L^p(\mathbb{R}^N)$ implies existence of a subsequence $h_n \to 0$ such that $D_{h_n}^l u \rightharpoonup v$ weakly in $L^p$, for some $v$, by the Banach—Alaoglu Theorem 2.31. (We proved this weak sequential compactness result only for $p = 2$, but it remains true in the "reflexive" range $1 < p < \infty$.)

We claim that $u_{x_l} = v$ weakly; indeed, for all test functions $\phi \in C_c^\infty(\mathbb{R}^N)$,

$$\begin{aligned} \int u\phi_{x_l}\,dx &= \lim_{n\to\infty} \int u(D^l_{-h_n}\phi)\,dx \\ &= -\lim_{n\to\infty}\int (D^l_{h_n}u)\phi\,dx \qquad \text{by (5.6)} \\ &= -\int v\phi\,dx \end{aligned}$$

by weak convergence. Therefore $u_{x_l} = v \in L^p$, and

$$\|u_{x_l}\|_{L^p} = \|v\|_{L^p} \le \liminf_{n\to\infty} \|D^l_h u\|_{L^p} \le K$$

with the first inequality holding by Exercise 2.20 (or the analogous result when $1 < p < \infty$). □

We iterate the $H^2$-regularity result to get higher regularity.

**Theorem 5.8** (Higher interior regularity)**.** *Let* $m \ge 0$*. Assume* $a^{jk} \in C^{m+1} \cap L^\infty(U)$ *and* $b^j, c \in C^m \cap L^\infty(U)$*, and* $f \in H^m(U)$*.*

*If* $u \in H^1(U)$ *solves* $Lu = f$ *weakly then* $u \in H^{m+2}_{loc}(U)$*. Further, on each open set* $V \Subset U$ *we have an* $H^{m+2}$*-norm estimate*

$$\|u\|_{H^{m+2}(V)} \le C(\|f\|_{H^m(U)} + \|u\|_{L^2(U)})$$

*where* $C = C(m, U, V, L)$*.*

The moral of the theorem is that the solution $u$ possesses two more derivatives than the data $f$ — which is desirable since $L$ is a second order differential operator!

*Proof for special case* $L = -\Delta$ *on* $\mathbb{R}^N$*. (This method extends to the general case [Evans].)* We proceed by induction on $m$. The theorem is true when $m = 0$, by Theorem 5.6.

Assume the theorem holds for $m - 1$. We prove it for $m$.

The idea is simply to differentiate the PDE. Formally we have $-\Delta u = f$ and differentiating with $D^\beta$ gives $-\Delta(D^\beta u) = D^\beta f$. Thus if $D^\beta f \in L^2$ then $D^\beta u \in H^2$ by Theorem 5.6, so that $u \in H^{|\beta|+2}$.

To make this idea rigorous, let $\beta$ be a multiindex of order $|\beta| = m$. By the induction hypothesis with $m - 1$ we know $u \in H^{m+1}_{loc}(U)$, with a norm estimate

$$\|u\|_{H^{m+1}(W)} \le C(\|f\|_{H^{m-1}(U)} + \|u\|_{L^2(U)})$$

whenever $V \Subset W \Subset U$. Take $\widetilde{v} \in C_c^\infty(W)$, and define $v = D^\beta \widetilde{v}$. Then because $Lu = f$ weakly we have

$$\begin{aligned} \alpha(u,v) &= \langle f, v\rangle_{L^2} \\ \int_W \nabla u \cdot \nabla v \, dx &= \int_W f v \, dx \\ \int_W \nabla(D^\beta u)\cdot \nabla \widetilde{v}\, dx &= \int_W (D^\beta f)\widetilde{v}\, dx \qquad \text{by integration by parts} \\ \alpha(D^\beta u, \widetilde{v}) &= \langle D^\beta f, \widetilde{v}\rangle_{L^2} \end{aligned}$$

where the left side makes sense because $u \in H^{m+1}(W)$ and the right side makes sense because $f \in H^m(W)$. This last formula extends to all $\widetilde{v} \in H_0^1(W)$, by density, and so $L(D^\beta u) = D^\beta f$ weakly in $W$. Thus $D^\beta u \in H^2_{loc}(W)$ by Theorem 5.6, and in particular $D^\beta u \in H^2(V)$ since $V \Subset W$. Hence $u \in H^{m+2}(V)$, with a suitable norm estimate. □

Letting $m \to \infty$ gives:

**Theorem 5.9** (**Infinite interior regularity**). *Assume* $a^{jk}, b^j, c \in C^\infty(U)$, *and* $f \in C^\infty(U)$. *If* $u \in H^1(U)$ *solves* $Lu = f$ *weakly, then* $u \in C^\infty(U)$ *and hence* $Lu = f$ *classically.*

*Proof.* Let $V \Subset U$, so that $a^{jk}, c, f$ are all smooth and bounded on $\overline{V}$. By applying the higher regularity result (Theorem 5.8) on $V$, we conclude $u \in H^{m+2}_{loc}(V)$ for each $m$. Since $V$ was arbitrary, we see $u \in H^{m+2}_{loc}(U)$ for each $m$.

Hence by the general Sobolev inequality (Theorem 3.23), we have $u \in C^j(V)$ for each $j \geq 1$ and all $V \Subset U$ having $C^1$-boundary (in particular, for all balls $V$ contained in $U$). Therefore $u \in C^\infty(U)$. □

## Boundary regularity

Next we prove regularity all the way up to the boundary, not just on compactly contained subdomains.

**Theorem 5.10** (Boundary $H^2$-regularity). *Assume* $a^{jk} \in C^1(\overline{U})$ *and* $b^j, c \in L^\infty(U)$, *and* $f \in L^2(U)$. *Suppose* $\partial U$ *is* $C^2$*-smooth.*

*If* $u \in H_0^1(U)$ *solves* $Lu = f$ *weakly then* $u \in H^2(U)$, *with estimate*

$$\|u\|_{H^2(U)} \leq C(\|f\|_{L^2(U)} + \|u\|_{L^2(U)}) \tag{5.7}$$

*where* $C = C(U, L)$.

The differences between this theorem and the interior result in Theorem 5.6 are that here we assume a Dirichlet boundary condition on $u$ (since $u$ belongs to $H^1_0$, not just to $H^1$) and we conclude that $u \in H^2$ (not just $u \in H^2_{loc}$).

*Note.* If $L$ is symmetric and all its eigenvalues are nonzero, then $\|u\|_{H^1(U)} \leq C\|f\|_{L^2(U)}$ by Theorem 5.1, in which case we deduce from (5.7) that $\|u\|_{H^2(U)} \leq C\|f\|_{L^2(U)}$. On the other hand, if $L$ has a zero eigenvalue (so that $Lu = 0 = f$ has a nontrivial solution), then the term $\|u\|_{L^2(U)}$ cannot be eliminated from (5.7). An example with zero eigenvalue is $L = -\frac{d^2}{dx^2} - 1$ on the interval $U = (0, \pi)$, for which we note $L(\sin x) = 0$.

*Proof sketch.* Since we already know $H^2$-regularity in the interior of the domain, the problem reduces to estimating $u$ near a point on the boundary. The boundary can be straightened on each such neighborhood, thereby reducing to the case of flat boundary $\{x_N = 0\}$. Then the proof of the interior $H^2$-regularity result can be repeated (see Theorem 5.6), giving that $u_{x_k x_l} \in L^2$ all the way up to the boundary, except the argument cannot be applied when $l = N$ because in that case the point $x \pm he_N$ in the difference quotient might push across the boundary.

Thus the second derivatives of $u$ are all globally in $L^2$, except perhaps for $u_{x_N x_N}$. To prove square integrability of that last derivative we call on the PDE, which rearranges to say

$$u_{x_N x_N} = \frac{1}{a^{NN}}\Big( - \sum_{(j,k)\neq(N,N)} a^{jk} u_{x_j x_k} - \sum_{j,k} (a^{jk})_{x_k} u_{x_j} + \sum_j b^j u_{x_j} + cu - f \Big) \in L^2.$$

See [Evans] for the proof of the full theorem. □

Iterating the $H^2$-regularity result (Theorem 5.10) yields higher regularity, and one arrives in the limit at:

**Theorem 5.11 (Infinite regularity up to the boundary).** *Assume* $a^{jk}, b^j, c \in C^\infty(\overline{U})$*, and* $f \in C^\infty(\overline{U})$*. Suppose* $\partial U$ *is* $C^\infty$*-smooth.*

*If* $u \in H^1_0(U)$ *solves* $Lu = f$ *weakly, then* $u \in C^\infty(\overline{U})$ *and hence* $Lu = f$ *classically with* $u = 0$ *on the boundary.*

The proof can be found in [Evans].

## Regularity theory in 1 dimension

The 1-dimensional case can be handled directly, as the next two exercises illustrate for the special case of Poisson's equation.

***Exercise*** 5.3 (Regularity for Poisson's equation in 1 dimension). Fix $f \in L^2(a, b)$ where $-\infty < a < b < \infty$.
(i) Suppose $u \in H^1(a, b)$ satisfies the ODE $-u'' = f$ in the weak sense, which means

$$\int_a^b u'v' \, dx = \int_a^b fv \, dx$$

for all $v \in H_0^1(a, b)$. Show that $u$ possesses two weak derivatives, and that the second weak derivative satisfies $-u'' = f$. Hence show $u \in H^2(a, b) \cap C^1[a, b]$.
(ii) Prove that if also $f \in C[a, b]$, then $u \in C^2[a, b]$ and so $-u'' = f$ classically.

***Exercise*** 5.4 (Regularity for Dirichlet eigenfunctions). Suppose $u \in H_0^1(0, \pi)$ is a weak eigenfunction of the second derivative operator, meaning $-u'' = \lambda u$ weakly for some $\lambda \in \mathbb{R}$, and $u \not\equiv 0$.

Prove that $-u'' = \lambda u$ classically. Hence by classical ODE theory and the boundary conditions $u(0) = 0 = u(\pi)$, we have $u(x) = c \sin(nx)$ for some constant $c \neq 0$ and some $n \in \mathbb{N}$. The eigenvalue is $\lambda = n^2$.

## 5.3 Weak maximum principles

The physical motivation for the maximum principle is that if $u(x)$ represents a steady state temperature, with $-\Delta u = 0$, then the maximum value of $u$ should occur on the boundary, because "heat flows from hot to cold" and so an interior point could never be the hottest.

Mathematically, we will see that maximum principle methods provide pointwise information about solutions, as opposed to the integral norm information gained so far from Sobolev spaces and their associated "energy methods".

The tools behind the maximum principle are familiar from calculus: if $u(x)$ is maximal at $x^*$ then

$$u_{x_j}(x^*) = 0 \tag{5.8}$$

for each $j$, by the first derivative test from calculus, and also

$$\frac{d^2}{dt^2} u(x^* + ty)\Big|_{t=0} \leq 0 \tag{5.9}$$

for each vector $y$, by the second derivative test. This last condition evaluates (by the chain rule) to say that

$$\sum_{j,k} u_{x_j x_k}(x^*) y_j y_k \leq 0,$$

which can be written $y^T(D^2u(x^*))y \leq 0$ where $D^2u$ is the Hessian matrix.

For the remainder of this chapter we assume $L$ is not in divergence form as previously assumed, but is in **nondivergence form**

$$\boxed{Lu = -\sum_{j,k} a^{jk}u_{x_jx_k} + \sum_j b^j u_{x_j} + cu} \tag{5.10}$$

where $a^{jk}, b^j, c \in C(\overline{U})$ and the $a^{jk}$ are uniformly elliptic and symmetric ($a^{jk} = a^{kj}$).

**Theorem 5.12** (Weak maximum principle for $c \equiv 0$)**.** *Suppose* $u \in C^2(U) \cap C(\overline{U})$, *and* $c \equiv 0$.
*(i) If* $Lu \leq 0$ *in* $U$ *then*

$$\max_{\overline{U}} u = \max_{\partial U} u.$$

*(ii) If* $Lu \geq 0$ *in* $U$ *then*

$$\min_{\overline{U}} u = \min_{\partial U} u.$$

*(iii) If* $Lu = 0$ *in* $U$ *then*

$$\min_{\partial U} u \leq u(x) \leq \max_{\partial U} u, \qquad x \in U.$$

In other words, subsolutions achieve their maximum on the boundary, while supersolutions achieve their minimum on the boundary.

An illustrative example in one dimension is the second derivative operator $L = -\frac{d^2}{dx^2}$. If $Lu \leq 0$ then $u'' \geq 0$, which means that $u$ is convex. A convex function on an interval always attains its maximum value at an endpoint, as one can understand by sketching some typical graphs.

*Proof.* Part (i).
Step 1 — The case $Lu < 0$. Start by making the stronger assumption that $Lu < 0$ in $U$. The continuous function $u$ attains its maximum at some point $x^* \in \overline{U}$. Suppose $x^* \in U$. We will deduce $Lu(x^*) \geq 0$, a contradiction, so that $x^*$ must in fact belong to $\partial U$, as we want to show.

Clearly $u_{x_j}(x^*) = 0$ by (5.8), and of course $c \equiv 0$ by hypothesis. To estimate the remaining terms in $Lu(x^*)$, we orthogonally diagonalize the matrix $A(x^*) = (a^{jk}(x^*))$ as

$$A(x^*) = M \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_N \end{pmatrix} M^T$$

where the eigenvalues along the diagonal satisfy $d_1, \dots, d_N \geq \theta > 0$. We have

$$\begin{aligned}
\sum_{j,k} a^{jk}(x^*) u_{x_j x_k}(x^*) &= \sum_{j,k,l} M_{jl} d_l M_{kl}\, u_{x_j x_k}(x^*) \\
&= \sum_l d_l \frac{d^2}{dt^2} u(x^* + t y^l)\Big|_{t=0} \\
&\qquad \text{where } y^l = \text{the } l\text{-th column of } M \\
&\leq 0 \qquad \text{by the second derivative test (5.9).}
\end{aligned}$$

Hence $Lu(x^*) \geq 0$, giving the desired contradiction.

Step 2 — The case $Lu \leq 0$. Now suppose only $Lu \leq 0$ in $U$. Fix a constant $\delta > \|b^1\|_\infty/\theta$. Let $\varepsilon > 0$ and define $\widetilde{u} = u + \varepsilon \exp(\delta x_1)$. Then

$$\begin{aligned}
L\widetilde{u} &= Lu + \varepsilon \exp(\delta x_1)(-a^{11}\delta^2 + b^1 \delta) \\
&\leq 0 + \varepsilon \exp(\delta x_1)(-\theta\delta + \|b^1\|_{L^\infty})\delta \qquad \text{since } a^{11} \geq \theta \text{ by ellipticity} \\
&< 0
\end{aligned}$$

by choice of the constant $\delta$. Hence $\widetilde{u}$ attains its maximum on $\partial U$, by Step 1. Letting $\varepsilon \to 0$ completes the proof, since

$$\begin{aligned}
\max_{\partial U} u \leq \max_{\overline{U}} u &\leq \max_{\overline{U}} \widetilde{u} \\
&= \max_{\partial U} \widetilde{u} \qquad \text{by Step 1} \\
&\leq \max_{\partial U} u + \varepsilon \max_{\partial U} \exp(\delta x_1) \\
&\to \max_{\partial U} u
\end{aligned}$$

as $\varepsilon \to 0$.

Part (ii). Apply part (i) to $-u$.

Part (iii). Combine parts (i) and (ii). □

**Corollary 5.13** (Pointwise continuous dependence on the boundary data)**.** *Assume* $c \equiv 0$*, and that the functions* $u, v \in C^2(U) \cap C(\overline{U})$ *satisfy* $Lu = Lv$*. Then*

$$\|u - v\|_{L^\infty(U)} \leq \|u - v\|_{L^\infty(\partial U)}.$$

*Proof.* Apply the weak maximum principle Theorem 5.12(iii) to $u - v$. □

***Exercise*** 5.5 (Interior control on the gradient). Assume $Lu = 0$ in $U$, where $u \in C^3(\overline{U})$ and $Lu = -\sum_{j,k} a^{jk} u_{x_j x_k}$ with $a^{jk} \in C^1(\overline{U})$ (we assume $b^j = 0, c = 0$). Let $v = |\nabla u|^2 + \sigma u^2$.

(i) Show that if the constant $\sigma$ is chosen large enough, then $L\nu \leq 0$.
*Hint.* Use somewhere that $(Lu)_{x_l} = 0$. Your value of $\sigma$ must be independent of $u$.
(ii) Deduce that

$$\|\nabla u\|_{L^\infty(U)} \leq C\big(\|u\|_{L^\infty(\partial U)} + \|\nabla u\|_{L^\infty(\partial U)}\big)$$

where $C = C(L, U)$ is independent of $u$.

*Note.* Thus the gradient in the interior is controlled by the values of the function and its gradient on the boundary. For the special case of harmonic functions ($-\Delta u = 0$) we do not need the boundary norm of $u$ on the right side, since the derivatives of a harmonic function are again harmonic and hence the maximum principle applies directly to $u_{x_j}$ for each $j$.

**Theorem 5.14** (Weak maximum principle for $c \geq 0$)**.** *Suppose* $u \in C^2(U) \cap C(\overline{U})$, *and* $c \geq 0$ *in* $U$. *If* $Lu = 0$ *in* $U$ *then*

$$\max_{\overline{U}} |u| = \max_{\partial U} |u|.$$

In other words, solutions achieve their maximum magnitude on the boundary. This result fails when $c < 0$. For example, in one dimension the function $u = \sin x$ satisfies $Lu = 0$ for the operator $L = -\frac{d^2}{dx^2} - 1$ on the interval $U = (0, \pi)$, but obviously $|u|$ does not attain its maximum at the endpoints of the interval. Essentially, the problem when $c < 0$ is that $u$ might be an eigenfunction of the purely second order part of the operator.

*Proof.* Let $V = \{x \in U : u(x) > 0\}$ and $W = \{x \in U : u(x) < 0\}$ be the open sets where $u$ is positive and negative, respectively.

Suppose $V$ is nonempty. Then $(L - c)u = -cu \leq 0$ on $V$, since $c \geq 0$ and $u > 0$ on $V$. Hence

$$u(x) \leq \max_{y \in \partial V} u(y), \qquad x \in V,$$

by the weak maximum principle Theorem 5.12(i) applied to the operator $L - c$.

Obviously $\partial V \subset \overline{U}$. Thus for each $y \in \partial V$, either $y \in \partial U$ or else $y \in U \cap \partial V$ (in which case $u(y) = 0$). Hence $\max_{y \in \partial V} u(y) \leq \max_{\partial U} |u|$, and so

$$0 < u(x) \leq \max_{\partial U} |u|, \qquad x \in V.$$

Applying the same argument to $-u$ on $W$ gives

$$0 < -u(x) \leq \max_{\partial U} |u|, \qquad x \in W,$$

and so we conclude $|u(x)| \leq \max_{\partial U} |u|$ for all $x \in U$. □

**Corollary 5.15** (Positivity of eigenvalues, via the maximum principle)**.** *Assume* $u \in C^2(U) \cap C(\overline{U})$ *and* $c \equiv 0$ *in* $U$*. If* $u$ *is a classical Dirichlet eigenfunction, meaning* $Lu = \lambda u$ *in* $U$ *and* $u = 0$ *on* $\partial U$*, then* $\lambda > 0$*.*

***Exercise*** 5.6. Prove Corollary 5.15 by using the maximum principle.

## 5.4 Strong maximum principles

We will show $u$ cannot attain an interior maximum (except when $u$ is constant), so that the maximum is attained *only* on the boundary. We start with a powerful technical result that says the normal derivative must be positive at a boundary maximum point.

Continue to assume the operator $L$ has nondivergence form, as in (5.10).

**Lemma 5.16** (Hopf's lemma for $c \equiv 0$)**.** *Assume* $u \in C^2(U) \cap C^1(\overline{U})$*, and* $c \equiv 0$ *in* $U$*. Suppose* $Lu \le 0$ *in* $U$ *and that* $x^* \in \partial U$ *is a strict maximum point on the boundary, meaning*

$$u(x^*) > u(x), \qquad x \in U.$$

*Suppose further that some ball* $B$ *lying in* $U$ *has* $x^*$ *on its boundary (note that the existence of such a ball is automatic when* $\partial U$ *is* $C^2$*-smooth). Then the outward normal derivative of* $u$ *is positive at* $x^*$*:*

$$\frac{\partial u}{\partial n}(x^*) > 0.$$

*The same conclusion holds if* $c \ge 0$ *in* $U$ *and* $u(x^*) \ge 0$*.*

The lemma holds regardless of whether $U$ is connected, since one may restrict to working solely on the ball $B$.

*Proof.*
Step 1— The perturbing function. By translation we may suppose the ball is centered at the origin, so that $B = B(0,r)$ for some $r > 0$. Let

$$\begin{aligned} v(x) &= \exp(-\beta|x|^2) - \exp(-\beta r^2) \\ &= \left[\exp(-\beta x_1^2) \cdots \exp(-\beta x_N^2)\right] - \exp(-\beta r^2) \end{aligned}$$

where $\beta = \beta(L,r) > 0$ is a constant to be chosen later. We compute

$$\begin{aligned} Lv &= \exp(-\beta|x|^2)\Big\{\sum_{j,k} a^{jk}(-4\beta^2 x_j x_k + 2\beta\delta_{jk}) - \sum_j b^j 2\beta x_j\Big\} \\ &\le \exp(-\beta|x|^2)\Big\{-4\beta^2\theta|x|^2 + 2\beta\sum_j a^{jj} + 2\beta|b||x|\Big\} \end{aligned}$$

by ellipticity. Hence

$$Lv \leq \exp(-\beta|x|^2)\big(-\beta^2\theta r^2 + 2\beta C + 2\beta C r\big) \qquad \text{whenever } \frac{r}{2} < |x| < r,$$

where $C = C(L)$. By choosing $\beta > 0$ to be large enough we can insure

$$Lv < 0 \qquad \text{whenever } \frac{r}{2} < |x| < r.$$

Next choose $\varepsilon > 0$ small enough that $u(x) + \varepsilon v(x) < u(x^*)$ when $|x| = r/2$, using here that $x^*$ is a strict local maximum point for $u$. Further notice $u(x) + \varepsilon v(x) \leq u(x^*)$ when $|x| = r$, since $v(x) = 0$ when $|x| = r$.

Step 2— The weak maximum principle. Let $\widetilde{u} = u + \varepsilon v$, so that $L\widetilde{u} = Lu + \varepsilon Lv < 0$ in the annulus $V = \{x : \frac{r}{2} < |x| < r\}$, with $\widetilde{u} \leq u(x^*)$ on $\partial V$. Hence by the weak maximum principle (Theorem 5.12), we conclude $\widetilde{u} \leq u(x^*)$ in $V$, with equality at $x^* \in \partial V$. Hence

$$\begin{aligned} 0 &\leq \frac{\partial \widetilde{u}}{\partial n}(x^*) \\ &= \frac{\partial u}{\partial n}(x^*) - 2\varepsilon\beta r \exp(-\beta r^2), \end{aligned}$$

so that $\partial u/\partial n$ must be positive at $x^*$.

Step 3 — The case where $c \geq 0$ and $u(x^*) \geq 0$ is treated as above, except with additional terms involving $c$. We omit the verification. □

***Exercise*** 5.7 (Boundary behavior of eigenfunctions). Assume $u \in C^2(U) \cap C^1(\overline{U})$, $c \equiv 0$ in $U$, and that $U$ has $C^2$-smooth boundary. Suppose $u$ is a classical Dirichlet eigenfunction of $L$, with $Lu = \lambda u$ in $U$ and $u = 0$ on $\partial U$.

Show that if $u$ is negative near a boundary point $x^* \in \partial U$ (meaning $u < 0$ on $U \cap B(x^*, r)$, for some $r$), then the outward normal derivative of $u$ is positive at $x^*$. (Correspondingly, the normal derivative is negative if $u > 0$ on $U \cap B(x^*, r)$.)

Now we strengthen the maximum principle, showing that the only way a subsolution can attain its maximum at an interior point is for the function to be constant.

**Theorem 5.17** (Strong maximum principle for $c \equiv 0$). *Suppose* $u \in C^2(U) \cap C(\overline{U})$, *and* $c \equiv 0$.
*(i) If* $Lu \leq 0$ *in* $U$ *and* $u(x) = \max_{\overline{U}} u$ *for some* $x \in U$, *then* $u \equiv$ *constant.*
*(ii) If* $Lu \geq 0$ *in* $U$ *and* $u(x) = \min_{\overline{U}} u$ *for some* $x \in U$, *then* $u \equiv$ *constant.*

Connectedness of the domain is essential to this result, since otherwise $u$ might be constant on one component (where it attains the maximum value) and nonconstant on another component.

*Proof.* Part (i). Consider the open set

$$V = \{x \in U : u(x) < \max_{\overline{U}} u\}$$

and its complement

$$K = \{x \in U : u(x) = \max_{\overline{U}} u\}.$$

Since $K$ is nonempty by hypothesis, we may choose $z \in K$.

Suppose $V$ is nonempty, and choose $y \in V$. Join $y$ to $z$ by a path in $U$, relying here on connectedness of the domain. Let $\delta > 0$ be less than the distance between this path and $\partial U$, and also small enough that the ball of radius $\delta$ centered at $y$ is contained entirely in the open set $V$. By considering the family of open balls of radius $\delta$ centered at points along the path, we see that as the center of the ball moves from $y$ to $z$, at some stage the boundary of the ball will first touch $K$. Let $x^* \in K$ denote a point of first touching, and call the ball $B$. Notice $B \subset V$, so that $u(x) < u(x^*)$ for all $x \in B$.

Applying Hopf's Lemma 5.16 on $B$ shows that $\frac{\partial u}{\partial n} > 0$ at $x^*$. On the other hand, $x^*$ is an interior maximum point of $u$ and so $\nabla u(x^*) = 0$. This contradiction shows that $V$ must be empty. Hence $u(x) = \max_{\overline{U}} u$ for all $x$ and so $u$ is constant, as we wanted to prove.

Part (ii). Apply Part (i) to $-u$. □

**Theorem 5.18** (Strong maximum principle for $c \geq 0$)**.** *Suppose* $u \in C^2(U) \cap C(\overline{U})$, *and* $c \geq 0$ *in* $U$.

*If* $Lu = 0$ *in* $U$ *and* $|u(x)| = \max_{\overline{U}} |u|$ *for some* $x \in U$, *then* $u \equiv$ *constant.*

***Exercise*** 5.8. Prove Theorem 5.18.

# Chapter 6

# Second order parabolic PDEs: existence, uniqueness, and maximum principles

**Reference** [Evans, Chapter 7].

Parabolic equations are time-evolution PDEs of the form

$$u_t + Lu = f.$$

The classic example is the diffusion or heat equation $u_t - \Delta u = f$, in which $u$ represents temperature and $f$ is a source–sink term representing creation or removal of heat.

## 6.1 Definitions

Fix the **terminal time** $T > 0$. Given the bounded domain $U$ we write

$$U_T = U \times (0, T]$$

for the **parabolic cylinder**, so that $(x, t) \in U_T$ means $x \in U$ and $t \in (0, T]$. We aim to find a solution $u : \overline{U_T} \to \mathbb{R}$ satisfying the PDE, Dirichlet boundary condition and initial condition:

$$\begin{aligned} u_t + Lu &= f \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0, T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}. && \text{IC} \end{aligned} \tag{6.1}$$

Here $L$ is a time-dependent operator in divergence form that is elliptic at each time:

$$\boxed{Lu = -\sum_{j,k} \left(a^{jk}(x,t)u_{x_j}\right)_{x_k} + \sum_j b^j(x,t)u_{x_j} + c(x,t)u}.$$

We assume for simplicity that the ellipticity constant $\theta$ is independent of $t$, that $a^{jk} = a^{kj}$, and that

$$a^{jk}, b^j, c \in L^\infty(U^T), \qquad f \in L^2(U_T), \qquad g \in L^2(U).$$

Define a sesquilinear form for each $t$ by

$$\boxed{\alpha(u,v;t) = \int_U \big(\sum_{j,k} a^{jk}u_{x_j}v_{x_k} + \sum_j b^j u_{x_j} v + cuv\big)\,dx}. \tag{6.2}$$

*Remark* 6.1. A nonzero Dirichlet boundary condition can be handled by subtraction, as follows. Assume $h(x,t)$ is defined for all $(x,t) \in \overline{U_T}$ (not just for $x \in \partial U$). To solve the parabolic PDE with nonzero BC $u = h$ on $\partial U \times [0,T]$, we let $\widetilde{u} = u - h$ and solve the corresponding problem for $\widetilde{u}$:

$$\begin{aligned} \widetilde{u}_t + L\widetilde{u} = \widetilde{f} &\stackrel{\text{def}}{=} f - h_t - Lh \quad \text{in } U_T, && \text{PDE} \\ \widetilde{u} &= 0 \quad \text{on } \partial U \times [0,T], && \text{BC} \\ \widetilde{u} = \widetilde{g} &\stackrel{\text{def}}{=} g - h(\cdot, 0) \quad \text{on } U \times \{0\}. && \text{IC} \end{aligned}$$

This new problem has zero BC and so fits the form of (6.1). We would simply need suitable hypotheses on $h$ to insure $\widetilde{f} \in L^2(U_T)$ and $\widetilde{g} \in L^2(U)$.

## 6.2 Galerkin approximate solutions

We will solve an approximate problem on a finite dimensional subspace of functions, and then pass to a limit. The basis functions for the subspaces will be eigenfunctions of the Laplacian. That might seem strange, because the operator $L$ is not the Laplacian. We would rather use eigenfunctions of $L$, but $L$ is nonsymmetric and so might not have eigenfunctions. Also, the coefficients of $L$ depend on time and so any eigenfunctions would depend on $t$, which would interfere with the "separation of variables" solution we are aiming to construct. To avoid such obstacles, we work with eigenfunctions of the fixed operator $-\Delta$.

Assume $\{w_k\} \subset H^1_0(U)$ is an ONB of weak eigenfunctions of $-\Delta$ on $U$. Fix $m \geq 1$. Suppose we are given real-valued, absolutely continuous coefficient functions

$d_m^k(t)$ and define a function

$$u_m(t) = \sum_{k=1}^{m} d_m^k(t) w_k, \qquad t \in [0,T],$$

where we suppress the $x$-dependence on both sides of the equation. (To show the $x$-dependence, we could write $w_k(x)$ on the right side and $u_m(t)(x)$ on the left side.) Since $w_k \in H_0^1(U)$ we have $u_m(t) \in H_0^1(U)$ for each $t \in [0,T]$.

We want $u_m$ to satisfy the **weak projected form** of the parabolic problem, which consists of the weak projected PDE

$$\langle (u_m)', w_k \rangle_{L^2} + a(u_m, w_k; t) = \langle f, w_k \rangle_{L^2}, \qquad k = 1, \ldots, m, \tag{6.3}$$

for almost every $t \in (0,T]$, along with the projected initial condition that

$$u_m(0) = L^2\text{-projection of } g \text{ onto the span of } \{w_1, \ldots, w_m\}.$$

This initial condition says

$$d_m^k(0) = \langle g, w_k \rangle_{L^2}, \qquad k = 1, \ldots, m.$$

Note the Dirichlet boundary condition holds automatically in the trace sense, because $u_m(t) \in H_0^1(U)$.

*Note.* The weak form of the PDE is motivated (as usual) by multiplying the equation $u_t + Lu = f$ by $w_k$ and formally integrating by parts.

**Theorem 6.2** (Construction of approximate solutions)**.** *A solution* $u_m$ *exists and is unique, for each* $m \geq 1$.

*Proof.* Notice that

$$(u_m)'(t) = \sum_{k=1}^{m} (d_m^k)'(t) w_k,$$

so that

$$\langle u_m'(t), w_k \rangle_{L^2} = (d_m^k)'(t).$$

And by linearity,

$$a(u_m, w_k; t) = \sum_{l=1}^{m} e^{kl}(t) d_m^l(t)$$

where $e^{kl}(t) = a(w_l, w_k; t)$. Clearly $e^{kl} \in L^\infty[0,T]$ since the coefficients in $L$ are bounded by hypothesis. Let $f^k(t) = \langle f, w_k \rangle_{L^2}$, so that $f^k \in L^2[0,T]$ because $f \in L^2(U_T)$.

Thus the projected weak form of the PDE (6.3) says

$$(d_m^k)'(t) + \sum_{l=1}^{m} e^{kl}(t) d_m^l(t) = f^k(t), \qquad t \in [0, T],$$

for $k = 1, \dots, m$, with initial condition $d_m^k(0) = \langle g, w_k \rangle_{L^2}$. Writing this equation in column vector form, with $\vec{d}_m = (d_m^1, \dots, d_m^m)^T$, $E = (e^{kl})$ and $\vec{f}_m = (f^1, \dots, f^m)^T$, we want to solve the linear system

$$\vec{d}_m' + E\vec{d}_m = \vec{f}_m, \qquad t \in [0, T],$$

with specified initial condition $\vec{d}_m(0)$. By ODE theory, the system has a unique solution $\vec{d}_m(t)$ that is absolutely continuous with respect to $t$ (and hence is differentiable a.e.). Or, you could prove existence of the solution by adapting the proof of Picard's Theorem 1.6 to apply to the space $X = C(J; \mathbb{R}^m)$ of continuous, vector-valued functions. □

## 6.3 Banach–space valued functions

Suppose $X$ is a separable Banach space with norm $\|\cdot\|_X$, and consider an $X$-valued function $u : [0, T] \to X$.

**Definition 6.3** ($L^p$ spaces)**.** For $1 \le p < \infty$ we define

$$L^p(0, T; X) = \{u : [0, T] \to X \text{ such that } \|u\|_{L^p(0,T;X)} < \infty\}$$

where

$$\|u\|_{L^p(0,T;X)} = \Big( \int_0^T \|u(t)\|_X^p \, dt \Big)^{1/p}.$$

Similarly

$$L^\infty(0, T; X) = \{u : [0, T] \to X \text{ such that } \|u\|_{L^\infty(0,T;X)} < \infty\}$$

where

$$\|u\|_{L^\infty(0,T;X)} = \operatorname{esssup}_{t \in [0,T]} \|u(t)\|_X.$$

These $L^p$-spaces are Banach spaces (proof omitted).

*Example* 6.4. If $X = H_0^1(U)$, then at each time $t$ we have a function $u(t)(\cdot) \in H_0^1$, which we regard as a function of $x$ and $t$ jointly by writing $u(x, t) = u(t)(x)$. The definition above gives

$$\|u\|_{L^2(0,T;H^1(U))}^2 = \int_0^T \int_U \big( |\nabla u|^2 + u^2 \big) \, dx \, dt.$$

Appendix E5 in [Evans] explains how to integrate X-valued functions. Intuitively, we regard the integral $\int_0^T u(t)\,dt$ as a Riemann sum. Such sums are well defined since linear combinations of elements of X belong to X.

**Definition 6.5** (Weak derivative)**.** A function $u \in L^1_{loc}(0,T;X)$ has **weak derivative** $v \in L^1_{loc}(0,T;X)$, written $u' = v$, if for all real-valued test functions $\phi \in C_c^\infty(0,T)$ we have

$$\int_0^T \phi'(t)u(t)\,dt = -\int_0^T \phi(t)v(t)\,dt.$$

(Here $\phi'$ and $\phi$ are real valued, while $u$ and $v$ are X-valued. Thus the integrals make sense.)

## 6.4 Energy estimates and weak solutions

To show that the sequence $\{u_m\}$ of weak approximate solutions converges to a solution of the original parabolic PDE, we will prove boundedness in suitable norms and then invoke weak compactness.

**Theorem 6.6** (Parabolic energy estimates)**.**

$$\|u_m\|_{L^\infty(0,T;L^2(U))} + \|u_m\|_{L^2(0,T;H^1(U))} + \|u_m'\|_{L^2(0,T;H^{-1}(U))} \le C\big(\|f\|_{L^2(U_T)} + \|g\|_{L^2(U)}\big)$$

*for all* $m \ge 1$*, for some constant* $C = C(T,U,L)$*.*

Here $H^{-1}$ denotes the dual of $H_0^1$, that is, the space of bounded linear functionals on $H_0^1$. In the theorem, we regard $u_m'(t) \in H_0^1$ as an element of $H^{-1}$ by associating with $u_m'(t)$ the bounded linear functional

$$F(v) = \langle u_m'(t), v\rangle_{L^2}, \qquad v \in H_0^1(U).$$

(You can check $|F(v)| \le C\|v\|_{H^1}$, so that the linear functional is bounded on $H_0^1$.) In other words, we use the $L^2$-pairing to obtain from any function in $H_0^1$ an element of $H^{-1}$. It might seem more natural to use the $H_0^1$-pairing, but the $L^2$-pairing arises naturally in the weak form of the parabolic PDE.

*Proof of Theorem 6.6.* Step 1 — energy dissipation inequality. The projected weak form of the PDE (6.3) implies that

$$\langle (u_m)', u_m\rangle_{L^2} + \alpha(u_m, u_m; t) = \langle f, u_m\rangle_{L^2} \tag{6.4}$$

since $u_m$ is a linear combination of $w_1, \dots, w_m$ at each time $t$. The second term can be estimated from below as follows, using ellipticity and Cauchy-with-$\varepsilon$:

$$\begin{aligned} \alpha(u_m, u_m; t) &\geq \theta \int_U |\nabla u_m|^2 \, dx - N \max_j \|b^j\|_{L^\infty} \int_U |\nabla u_m||u_m| \, dx - \|c\|_{L^\infty} \int_U u_m^2 \, dx \\ &\geq \frac{\theta}{2} \int_U |\nabla u_m|^2 \, dx - (\text{const.}) \int_U u_m^2 \, dx. \end{aligned}$$

After using this lower bound in (6.4), we deduce

$$\frac{d}{dt} \|u_m\|_{L^2}^2 + \theta \|u_m\|_{H^1}^2 \leq \|f\|_{L^2}^2 + C\|u_m\|_{L^2}^2. \tag{6.5}$$

Step 2 — Gronwall inequality argument. Write $\eta(t) = \|u_m\|_{L^2}^2$ and $\xi(t) = \|f\|_{L^2}^2$, where we recall that $u_m$ and $f$ both depend on $t$ as well as $x$. After discarding the $H^1$-norm from (6.5) we are left with a linear differential inequality for $\eta$:

$$\eta'(t) \leq C\eta(t) + \xi(t), \qquad t \in [0, T].$$

We multiply by the integrating factor $e^{-Ct}$ and integrate from $0$ to $t$:

$$\begin{aligned} \big(e^{-Ct}\eta(t)\big)' &\leq e^{-Ct}\xi(t) \leq \xi(t) \\ e^{-Ct}\eta(t) - \eta(0) &\leq \int_0^t \xi(s) \, ds \\ \eta(t) &\leq e^{Ct}\Big(\eta(0) + \int_0^t \xi(s) \, ds\Big) \\ \eta(t) &\leq e^{CT}\Big(\sum_{k=1}^m d_m^k(0)^2 + \int_0^T \xi(s) \, ds\Big). \end{aligned}$$

Hence

$$\|u_m(t)\|_{L^2}^2 \leq C\big(\|g\|_{L^2}^2 + \|f\|_{L^2(0,T;L^2(U))}^2\big) \tag{6.6}$$

where $C = C(T, L, U)$. Now we get the first part of the theorem by taking the supremum over $t$ on the left side.

Step 3 — The $H^1$-norm term. Integrating (6.5) from $0$ to $T$ gives that

$$\|u_m(T)\|_{L^2}^2 - \|g\|_{L^2}^2 + \theta \int_0^T \|u_m(t)\|_{H^1}^2 \, dt \leq \int_0^T \|f\|_{L^2}^2 \, dt + C \int_0^T \|u_m\|_{L^2}^2 \, dt.$$

The last term on the right can be estimated using (6.6), and so we deduce that

$$\|u_m\|_{L^2(0,T;H^1(U))}^2 \leq C\big(\|f\|_{L^2(0,T;L^2(U))}^2 + \|g\|_{L^2(U)}^2\big), \tag{6.7}$$

which is the second part of the theorem.

Step 4 — The $H^{-1}$-norm term. Consider an arbitrary element $v \in H_0^1(U)$, and decompose it as $v = w + w^\perp$ where $w = \sum_{k=1}^m c_k w_k$ and $w^\perp = \sum_{k=m+1}^\infty c_k w_k$, with convergence in $L^2$ and $H^1$. Then

$$\begin{aligned} \left|\langle u_m', v\rangle_{L^2}\right| &= \left|\langle u_m', w\rangle_{L^2}\right| \qquad \text{since } \langle u_m', w^\perp\rangle_{L^2} = 0 \\ &= \left|\langle f, w\rangle_{L^2} - \alpha(u_m, w; t)\right| \qquad \text{by the projected PDE (6.3)} \\ &\le C(\|f\|_{L^2} + \|u_m\|_{H^1})\|w\|_{H^1} \\ &\le C(\|f\|_{L^2} + \|u_m\|_{H^1})\|v\|_{H^1} \end{aligned}$$

since $w \perp w^\perp$ in $L^2$ and $\nabla w \perp \nabla w^\perp$ in $L^2$. Thus when $u_m'$ is considered as a bounded linear functional on $H_0^1$, acting via the $L^2$-pairing, it has norm

$$\|u_m'\|_{H^{-1}} \le C(\|f\|_{L^2} + \|u_m\|_{H^1}).$$

Squaring and integrating from $0$ to $T$ and then using (6.7) proves that

$$\|u_m'\|^2_{L^2(0,T;H^{-1}(U))} \le C\big(\|f\|^2_{L^2(0,T;L^2(U))} + \|g\|^2_{L^2}\big),$$

which is the third part of the theorem. □

Now we state the main existence theorem.

**Theorem 6.7** (Parabolic existence and uniqueness result)**.** *The parabolic problem* (6.1) *has a unique weak solution. That is, a unique function* $u \in L^2(0,T;H_0^1(U))$ *exists that has weak derivative* $u' \in L^2(0,T;H^{-1}(U))$ *such that*

$$\begin{aligned} \langle u', v\rangle + \alpha(u, v; t) &= \langle f, v\rangle_{L^2(U)}, \qquad \forall v \in H_0^1(U), \ \textit{almost every } t \in (0,T), \\ u(0) &= g. \end{aligned}$$

The statement of the theorem relies implicitly on the weakly differentiable mapping $u(t)$ being absolutely continuous with respect to $t$, and hence differentiable almost everywhere, the proof of which proceeds just as for real-valued functions.

*Proof sketch.*

Step 1 — Existence. By the energy estimates in Theorem 6.6, the sequence $\{u_m\}$ of approximate solutions is bounded in $L^2(0,T;H_0^1(U))$ and the sequence of $t$-derivatives $\{u_m'\}$ is bounded in $L^2(0,T;H^{-1}(U))$. Thus the Banach–Alaoglu Theorem 2.31 provides a weakly convergent subsequence, and the limiting function is the desired solution $u$; for the detailed proof, see [Evans, Chapter 7].

Step 2 — Uniqueness. The task is to show that if $u_t + Lu = 0$ weakly, with initial condition $u = 0$ when $t = 0$, then $u \equiv 0$. One does this by mimicking for $u$ the energy estimates and Gronwall argument that led to (6.6). □

One can obtain higher regularity results, when $f$ and $g$ have some smoothness. See [Evans].

***Exercise*** 6.1 (Parabolic decay rate when $b^j \equiv c \equiv 0$). Assume $u(x,t)$ is a smooth solution of

$$\begin{aligned} u_t + Lu &= 0 \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0,T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}, && \text{IC} \end{aligned}$$

where $Lu = -\sum_{j,k} \left(a^{jk}(x,t)u_{x_j}\right)_{x_k}$. That is, $u$ is a smooth function of $x$ and $t$, and it satisfies the PDE and initial condition and boundary conditions in the classical sense. The boundary $\partial U$ is assumed to be smooth also.

Prove the exponential $L^2$-decay estimate

$$\|u(\cdot,t)\|_{L^2(U)} \le e^{-\delta t}\|g\|_{L^2(U)}, \qquad t \ge 0,$$

for some constant $\delta > 0$, by multiplying the PDE with $u$ and integrating to obtain an energy estimate.

Intuitively, the Dirichlet boundary condition "sucks heat out of the domain exponentially fast."

*Remarks.* In the special case when $L$ is independent of $t$, exponential decay of the solution can be justified using eigenfunctions and eigenvalues, as follows. The solution has the form

$$u(x,t) = \sum_{k=1}^{\infty} e^{-\lambda_k t} d_k u_k(x) \tag{6.8}$$

where the $u_k$ are the Dirichlet eigenfunctions of $L$ with corresponding eigenvalues $\lambda_k$, and the $d_k$ are the coefficients of the initial data: $g = \sum_{k=1}^{\infty} d_k u_k$. (One can justify (6.8) formally by observing that $u_t + Lu = 0$ with $u(x,0) = g(x)$, and can justify it rigorously by differentiating the $k$-th coefficient $\langle u(\cdot,t), u_k\rangle_{L^2(U)}$ with respect to $t$ and solving the resulting ODE.) From (6.8) one computes that

$$\|u(\cdot,t)\|_{L^2(U)}^2 = \sum_{k=1}^{\infty} e^{-2\lambda_k t} d_k^2 \le e^{-2\lambda_1 t} \sum_{k=1}^{\infty} d_k^2 = e^{-2\lambda_1 t}\|g\|_{L^2(U)}^2,$$

which gives the desired exponential decay with $\delta = \lambda_1$.

Thus the point of the exercise is to treat the case where coefficients of $L$ can depend on $t$.

## 6.5 Maximum principles

Assume for the remainder of the chapter that $L$ has nondivergence form:

$$\boxed{Lu = -\sum_{j,k} a^{jk}(x,t)u_{x_jx_k} + \sum_j b^j(x,t)u_{x_j} + c(x,t)u}$$

where in addition to our earlier hypotheses we assume the $a^{jk}, b^j, c$ are continuous on $\overline{U^T} = \overline{U} \times [0,T]$. Define the **parabolic boundary** of the parabolic cylinder to be

$$\Gamma_T = \text{bottom} \cup \text{sides} = \big(U \times \{0\}\big) \cup \big(\partial U \times [0,T]\big).$$

Write $C_1^2$ for the class of functions having two continuous derivatives in the $x$-directions and $1$ continuous derivative in the $t$-direction.

**Theorem 6.8** (Weak parabolic maximum principle for $c \equiv 0$). *Suppose* $u \in C_1^2(U_T) \cap C(\overline{U_T})$, *and* $c \equiv 0$.
*(i) If* $u_t + Lu \leq 0$ *in* $U_T$ *then*

$$\max_{\overline{U_T}} u = \max_{\Gamma_T} u.$$

*(ii) If* $u_t + Lu \geq 0$ *in* $U_T$ *then*

$$\min_{\overline{U_T}} u = \min_{\Gamma_T} u.$$

*(iii) If* $u_t + Lu = 0$ *in* $U_T$ *then* $u$ *attains its maximum value either initially or at the spatial boundary, and similarly for the minimum value:*

$$\min_{\Gamma_T} u \leq u(x,t) \leq \max_{\Gamma_T} u, \qquad (x,t) \in U_T.$$

In other words, subsolutions achieve their maximum on the parabolic boundary, while supersolutions achieve their minimum on the parabolic boundary. The intuition in terms of heat flow is that the hottest point must occur either at the initial time or else on the boundary. The hottest point should not occur at an interior point at a positive time, because to create such a situation we would need heat to flow *towards* a hot point. (This claim is made precise in the strong maximum principle below.)

*Proof — In-class exercise.*
Step 1 — Assume $u_t + Lu < 0$ in $U_T$. Suppose $u$ attains its maximum at a point $(x^*, t^*) \in \overline{U_T}$.

Maximum in interior: show that if $x^* \in U$ and $0 < t^* < T$ then a contradiction ensues. Where does the argument use that $c \equiv 0$?
Maximum on top: show that if $x^* \in U$ and $t^* = T$ then a contradiction ensues.

Hence either $x^* \in \partial U$ or else $t^* = 0$, so that $u$ attains its maximum on the parabolic boundary.

Step 2 — Assume $u_t + Lu \leq 0$ in $U_T$. Let

$$\widetilde{u}(x,t) = u(x,t) - \varepsilon t$$

so that $\widetilde{u}_t + L\widetilde{u} = u_t + Lu - \varepsilon < 0$ in $U_T$. Thus $\widetilde{u}$ attains its maximum on the parabolic boundary. Letting $\varepsilon \to 0$ shows the same holds for $u$.

Step 3 — Parts (ii) and (iii). For part (ii), apply part (i) to $-u$. For part (iii), combine parts (i) and (ii). □

Now we sharpen the conclusion to say that if a solution attains its maximum value inside the domain, then the solution must be constant up to that time.

**Theorem 6.9** (Strong parabolic maximum principle for $c \equiv 0$)**.** *Suppose* $u \in C^2_1(U_T) \cap C(\overline{U_T})$*, and* $c \equiv 0$.
*(i) If* $u_t + Lu \leq 0$ *in* $U_T$ *and* $u$ *attains its maximum at* $(x^*, t^*) \in U_T$ *then* $u \equiv$ *constant for* $t \leq t^*$*.*
*(ii) If* $u_t + Lu \geq 0$ *in* $U_T$ *and* $u$ *attains its minimum at* $(x^*, t^*) \in U_T$ *then* $u \equiv$ *constant for* $t \leq t^*$*.*

This result relies on the parabolic Harnack principle, which says that solutions to parabolic equations cannot change too much from point to point. See [Evans].

***Exercise*** 6.2 (**Infinite propagation speed** for parabolic equations). Take $c \equiv 0$.
(i) Suppose $u \in C^2_1(U_T) \cap C(\overline{U_T})$, and that

$$u_t + Lu = 0$$

in $U_T$, with $u = 0$ on the boundary (at each $t$). Assume $u$ has compact support initially (at $t = 0$), with $u(x,0) \geq 0$ for all $x$ and $u(x_0, 0) > 0$ for some $x_0 \in U$. Prove that $u(x,t) > 0$ for all $x \in U$ and all $t > 0$.
(ii) Explain why part (i) can be interpreted as showing infinite propagation speed for the solution of the parabolic equation.

Infinite propagation speed is easily shown for the classical heat equation $u_t - \Delta u = 0$, by means of the fundamental solution formula in terms of the Gaussian kernel:

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{\mathbb{R}} \exp\big(-(x-y)^2/4t\big) g(y)\, dy$$

in $1$ dimension, where $g(x) = u(x, 0)$ is the initial condition. Indeed, if $g \geq 0$ has compact support, and $g > 0$ at some point, then the formula shows $u(x, t) > 0$ for all $x \in \mathbb{R}$ and all $t > 0$.

The heat equation illustrates also the **smoothing effect** of parabolic PDEs: the initial data $g$ can be nonsmooth (for example, with jumps), but the solution $u$ is a smooth function of $x$, for each $t > 0$. The smoothing effect holds for the general parabolic equation too, assuming smooth coefficients, but in this course we do not prove such "interior parabolic regularity" results.

To slow down the diffusion and get **finite** speed of propagation, one can consider the nonlinear **porous medium** equation

$$u_t = (u^p u_x)_x$$

where $p > 0$ is fixed. The fact that the diffusivity $u^p$ approaches $0$ as the solution approaches $0$ leads to the solution having *finite* speed of propagation; proof omitted.

# Chapter 7

# Second order hyperbolic PDEs: existence, uniqueness, and finite speed of propagation

**Reference** [Evans, Chapter 7].

Hyperbolic equations are time-evolution PDEs of the form

$$u_{tt} + Lu = f.$$

The classic example is the wave equation $u_{tt} - \Delta u = f$, in which $u$ represents (for example) transverse displacement of a membrane from equilibrium and $f$ is an external forcing term. Other wave phenomena modeled by the wave equation include electromagnetic radiation (light and radio waves), certain kinds of water wave, seismic waves, and sound waves. Quantum wavefunctions are closely related too, as they satisfy the Schrödinger equation.

## 7.1 Definitions

With notation as in the previous chapter, we aim to find a solution $u : \overline{U_T} \to \mathbb{R}$ satisfying the PDE, Dirichlet boundary condition, and initial condition on both $u$ and $u_t$:

$$\begin{aligned} u_{tt} + Lu &= f \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0, T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}, && \text{IC} \\ u_t &= h \quad \text{on } U \times \{0\}. && \text{IC} \end{aligned} \tag{7.1}$$

Again $L$ is a time-dependent operator in divergence form that is elliptic at each time:

$$Lu = -\sum_{j,k} \left(a^{jk}(x,t)u_{x_j}\right)_{x_k} + \sum_j b^j(x,t)u_{x_j} + c(x,t)u.$$

For simplicity we assume the ellipticity constant $\theta$ is independent of $t$, and that

$$a^{jk}, b^j, c \in C^\infty(\overline{U^T}), \qquad f \in L^2(U_T), \qquad g \in H_0^1(U), \qquad h \in L^2(U).$$

*Note.* A nonzero boundary condition can be handled by subtraction; cf. the parabolic case in the previous chapter.

**Definition 7.1.** Call $u$ a **weak solution** of the hyperbolic problem (7.1) if

$$\begin{aligned} u &\in L^2(0,T; H_0^1(U)) \\ u' &\in L^2(0,T; L^2(U)) \\ u'' &\in L^2(0,T; H^{-1}(U)) \end{aligned}$$

and $u$ satisfies

$$\begin{aligned} \langle u'', v\rangle + \alpha(u,v;t) &= \langle f, v\rangle_{L^2(U)}, \qquad \forall v \in H_0^1(U), \text{ almost every } t \in (0,T), \\ u(0) &= g, \\ u'(0) &= h. \end{aligned}$$

One needs to show absolute continuity of $u$ and $u'$ so that $u'(t)$ and $u''(t)$ exist for almost every $t$, and so that the initial conditions $u(0)$ and $u'(0)$ are well-defined. Also, one must interpret the pairing $\langle u'', v\rangle$ appropriately in terms of the $L^2$-inner product; see [Evans].

## Traveling wave example

Write $\tau(x) = (1-|x|)_+$ for the "triangular pulse" function. Then

$$u(x,t) = \tau(x-t)$$

is a right-moving traveling wave solution of the wave equation $u_{tt} - u_{xx} = 0$ in 1-dimension. Clearly at each time $u(\cdot,t) \in H_0^1$, but $u_t(x,t)$ has jumps (as a function of $x$) and so $u_t(\cdot,t) \notin H^1$, although at least we have $u_t(\cdot,t) \in L^2$. This example motivates the above definition of weak solution.

This traveling pulse example also reminds us that hyperbolic equations are not (in general) smoothing, because corners in our initial data can persist in the solution when $t > 0$. Nonetheless, dispersive PDEs such as the Schrödinger equation can exhibit a certain amount of smoothing, and this issue of dispersive smoothing is an active topic of research.

## 7.2 Energy estimates and weak solutions

We will derive energy estimates for the wave equation, since this case involves already the key points needed in the general hyperbolic case, which we omit. So suppose $u$ is a smooth solution of

$$\begin{aligned} u_{tt} - \Delta u &= f \quad \text{in } \mathbb{R}^N \times (0, T), && \text{PDE} \\ u &= g \quad \text{when } t = 0, && \text{IC} \\ u_t &= h \quad \text{when } t = 0. && \text{IC} \end{aligned}$$

Instead of a boundary condition, suppose $u$ and its derivatives approach $0$ rapidly as $|x| \to \infty$, at each fixed time $t$, so that we may discard boundary terms when integrating by parts in the argument below.

To begin with, multiply the PDE by $2u_t$ and integrate over $\mathbb{R}^N$:

$$\begin{aligned} 2\int u_t(u_{tt} - \Delta u)\,dx &= 2\int u_t f\,dx \\ 2\int (u_t u_{tt} + \nabla u_t \cdot \nabla u)\,dx &= 2\int u_t f\,dx \qquad \text{by parts} \\ \frac{d}{dt}\int (u_t^2 + |\nabla u|^2)\,dx &= 2\int u_t f\,dx \end{aligned}$$

The left side is interpreted as the derivative of the total energy, since $\frac{1}{2}\int u_t^2\,dx$ represents the kinetic energy (remember $u_t$ is velocity) and $\frac{1}{2}\int |\nabla u|^2\,dx$ represents potential energy.

If $f \equiv 0$ (no external forcing) then the right side of the last equation is zero and so energy is conserved. If $f \not\equiv 0$ then we bound $2u_t f$ on the right side by $u_t^2 + f^2$, so that

$$\begin{aligned} \frac{d}{dt}\int (u_t^2 + |\nabla u|^2)\,dx &\le \int (u_t^2 + f^2)\,dx \\ &\le \int (u_t^2 + |\nabla u|^2)\,dx + \int f^2\,dx. \end{aligned}$$

Then by a Gronwall argument (which in this case means to multiply by $e^{-t}$ and integrate from $0$ to $t$) we find

$$\int_{\mathbb{R}^N} (u_t^2 + |\nabla u|^2)\,dx \le e^T\Big(\int_{\mathbb{R}^N} (h^2 + |\nabla g|^2)\,dx + \int_0^T \int_{\mathbb{R}^N} f^2\,dxdt\Big),$$

for each $t \in [0, T]$.

The last formula says that at each time, the total energy is bounded by the initial total energy plus a contribution from the forcing. In particular, at each time we have $u \in H^1$ and $u_t \in L^2$, as needed in the definition of weak solutions. This energy bound helps us prove:

**Theorem 7.2** (Hyperbolic existence and uniqueness result)**.** *The hyperbolic problem* (7.1) *has a unique weak solution.*

*Proof sketch.* One constructs approximate solutions using the Galerkin method, as in the parabolic case. These approximate projected weak solutions are bounded in suitable function spaces by energy estimates such as proved above for the wave equation. Therefore some subsequence of the approximate solutions converges by the Banach–Alaoglu Theorem 2.31. One can show this limiting function is a weak solution of the full problem.

For uniqueness, suppose two solutions exist and denote their difference by $u$, so that

$$\begin{aligned} u_{tt} + Lu &= 0 \quad \text{in } U_T, \\ u &= 0 \quad \text{on } \partial U \times (0, T], \\ u &= 0 \quad \text{when } t = 0, \\ u_t &= 0 \quad \text{when } t = 0. \end{aligned}$$

Then $u_t \equiv 0 \equiv |\nabla u|$ by the energy estimate, since the initial energy is zero and there is no forcing term. Hence $u \equiv$ const., and the constant must be zero due to the Dirichlet boundary condition. □

## 7.3 Finite speed of propagation

Parabolic equations have infinite propagation speed, as we saw in the previous chapter. In this section we show that solutions to hyperbolic equations have finite propagation speed.

Let $U = \mathbb{R}^N$ for simplicity. Suppose the coefficients $a^{jk}(x)$ are smooth and independent of $t$, and that the symmetric elliptic operator has the divergence form

$$Lu = -\sum_{j,k} \left(a^{jk} u_{x_j}\right)_{x_k}$$

with no lower order terms. Fix a point $x_0 \in \mathbb{R}^N$ and assume $q(x)$ solves

$$\begin{aligned} \sum_{j,k} a^{jk} q_{x_j} q_{x_k} &= 1 \quad \text{when } x \neq x_0, \qquad (7.2) \\ q(x) &> 0 \quad \text{when } x \neq x_0, \\ q(x_0) &= 0, \end{aligned}$$

with $q$ being continuous everywhere and smooth except perhaps at $x_0$, and with $q(x) \to \infty$ as $|x| \to \infty$. (For example, when $L = -\Delta$, equation (7.2) says $|q(x)|^2 = 1$, which has solution $q(x) = |x - x_0|$.)

Fix a time $t_0 > 0$ and define a "solid cone"

$$\mathcal{C} = \{(x,t) : q(x) < t_0 - t\}$$

with vertex at $(x_0, t_0)$.

**Theorem 7.3** (Finite propagation speed for hyperbolic equation)**.** *Assume* $u$ *is a smooth solution of*

$$u_{tt} + Lu = 0$$

*for* $x \in \mathbb{R}^N, t \geq 0$. *If* $u \equiv u_t \equiv 0$ *inside the cone initially (on* $\mathcal{C} \cap \{t = 0\}$*) then* $u \equiv 0$ *on the whole cone* $\mathcal{C}$*.*

Thus initial disturbances outside the cone at $t = 0$ take at least time $t_0$ to reach the point $x_0$.

*Proof.* Write

$$\mathcal{C}_t = \{x : q(x) < t_0 - t\}$$

for the cross-section of the cone at time $t$. Define the **local energy** at time $t$ to be

$$e(t) = \frac{1}{2}\int_{\mathcal{C}_t} (u_t^2 + \sum_{j,k} a^{jk} u_{x_j} u_{x_k})\, dx,$$

where the first term represents kinetic energy and the second is potential energy. We aim to show this energy is dissipated. Its time derivative is

$$e'(t) = \int_{\mathcal{C}_t} (u_t u_{tt} + \sum_{j,k} a^{jk} u_{x_j} u_{t x_k})\, dx - \frac{1}{2}\int_{\partial\mathcal{C}_t} (u_t^2 + \sum_{j,k} a^{jk} u_{x_j} u_{x_k}) \frac{1}{|\nabla q|}\, dS \tag{7.3}$$

where the first term comes from differentiating through the integral, and the second from differentiating the domain of integration (with the assistance of the co-area formula [Evans, Appendix C]). The negative sign reflects the fact that the cone cross-section $\mathcal{C}_t$ gets smaller as $t$ increases. (Technically, $e(t)$ is absolutely continuous and the second term in the derivative exists only for almost every $t$, but that is good enough for our purposes.)

We estimate the first term in $e'(t)$ as follows:

$$\begin{aligned}
&\int_{\mathcal{C}_t} (u_t u_{tt} + \sum_{j,k} a^{jk} u_{x_j} u_{tx_k})\, dx \\
&= \int_{\mathcal{C}_t} u_t(u_{tt} + Lu)\, dx + \int_{\partial\mathcal{C}_t} u_t \sum_{j,k} a^{jk} u_{x_j} n_k\, dS \qquad \text{by parts w.r.t. } x_k \\
&\leq \int_{\partial\mathcal{C}_t} |u_t| \big( \sum_{j,k} a^{jk} u_{x_j} u_{x_k} \big)^{1/2} \big( \sum_{j,k} a^{jk} n_j n_k \big)^{1/2} dS
\end{aligned}$$

by the PDE and by applying Cauchy–Schwarz to the $a$-inner product. The level set $\partial\mathcal{C}_t$ of $q$ has unit outward normal $n = \nabla q / |\nabla q|$. Substituting this normal into the last formula and calling on equation (7.2) for $q$ shows that the first term in $e'(t)$ satisfies

$$\int_{\mathcal{C}_t} (u_t u_{tt} + \sum_{j,k} a^{jk} u_{x_j} u_{tx_k})\, dx \leq \int_{\partial\mathcal{C}_t} \frac{1}{2} \big( u_t^2 + \sum_{j,k} a^{jk} u_{x_j} u_{x_k} \big) \frac{1}{|\nabla q|}\, dS$$

where we used also the elementary estimate $ab \leq \frac{1}{2}(a^2 + b^2)$.

Hence from (7.3) we conclude $e'(t) \leq 0$. That is, the local energy is dissipated. The local energy vanishes initially, by hypothesis, and so $e(t) = 0$ for all $t \in [0, t_0]$. Hence $u_t \equiv 0$ and $\nabla u \equiv 0$ in the cone, so that $u$ is constant there, and therefore $u$ vanishes identically in $\mathcal{C}$. □

# Chapter 8

# Semigroup theory

**References** [Brezis, Chapter 7], [Evans, Section 7.4], [Pazy, Chapter 1]

Semigroup theory takes a "dynamical systems" approach to parabolic and hyperbolic PDEs. Rather than employing the Galerkin method, which is practical but inelegant, we work with the "solution flow."

## 8.1 Semigroups, generators and resolvents

### Semigroups

**Definition 8.1.** Let $X$ be a real Banach space and $S(t) : X \to X$ a bounded linear operator, for each $t \geq 0$. Call the family of operators $\{S(t)\}_{t \geq 0}$ a **semigroup** if

- $S(0) = I$, meaning $S(0)x = x$ for each $x \in X$,
- $S(t+s) = S(t) \circ S(s)$ for all $s, t \geq 0$, meaning $S(t+s)x = S(t)\big(S(s)x\big)$ for each $x \in X$,
- $t \mapsto S(t)x$ is a continuous function of $t \geq 0$ ("continuity of the solution curve"), for each initial point $x \in X$.

If in addition each operator in the family has norm at most $1$ (meaning $\|S(t)x\| \leq \|x\|$ for all $x \in X, t \geq 0$), then we call $\{S(t)\}_{t \geq 0}$ a **contraction** semigroup.

*Note.* The semigroup property (the second property in the definition) implies that $S(t)$ and $S(s)$ commute, since $s + t = t + s$.

*Example* 8.2.
(i) The identify semigroup has $S(t) = I$ for all $t$ (which is not interesting at all).

(ii) Our first interesting example is the exponential example $S(t)x = e^{-t}x$, which gives a contraction semigroup. The semigroup property here relies simply on the law of exponents:

$$e^{-(t+s)} = e^{-t}e^{-s}.$$

(iii) A generalization is:

***Exercise*** 8.1. Prove that the matrix exponential

$$S(t) = \exp(At)$$

gives a contraction semigroup on $X = \mathbb{R}^N$, when $A$ is an $N \times N$ real symmetric matrix all of whose eigenvalues are less than or equal to $0$. (Recall the matrix exponential is defined by the usual power series, and that the law of exponents holds for $\exp(A+B)$ provided $A$ and $B$ commute.)

The matrix exponential shows the connection between semigroups and differential equations, because the linear ODE initial value problem

$$\frac{dx}{dt} = Ax, \qquad x(0) = u$$

has solution $x(t) = \exp(At)u = S(t)u$. Notice the matrix $A$ can be obtained from the semigroup by differentiating at $t = 0$, with $A = \frac{d}{dt}S(t)|_{t=0}$.

We aim to perform a similar analysis with the matrix $A$ replaced by an elliptic partial differential operator of second order, and with $x(t)$ replaced by a function of space and time, in which case the ODE transforms into a parabolic PDE like the heat equation. One wrinkle in this approach is that $L^2$ functions are not necessarily differentiable, and so $Au$ will not make sense for all functions $u$, only for those $u$ in a Sobolev space such as $H^2$. Fortunately, Sobolev functions are dense in $L^2$, and that will turn out to be good enough.

***Exercise*** 8.2.

(a) Show that continuity of $S$ as a mapping from $[0,\infty)$ to the space of bounded linear operators $X \to X$ implies continuity of the solution curves (continuity of $t \mapsto S(t)x$ for each fixed $x \in X$).

(b) Give an example where norm continuity fails and yet the solution curves are still continuous. (*Hint.* Diffusion equation.)

## Generators

Our task is to develop semigroup theory for the case where the "generator" $A$ associated with the semigroup is densely defined.

**Definition 8.3** (Generator of the semigroup)**.** Given a semigroup $\{S(t)\}_{t\geq 0}$, we define

$$Ax = \frac{d}{dt}S(t)x\Big|_{t=0+} = \lim_{t\to 0+}\frac{S(t)x - x}{t} \in X,$$

and write $D(A)$ for the subspace of $x \in X$ for which this limit exists. Call $A$ the (infinitesimal) **generator** of the semigroup.

In other words, $Ax$ is the tangent vector to the solution curve through $x$.

Our first theorem shows that the semigroup solves the ODE

$$\frac{dx}{dt} = Ax$$

provided the initial condition $x(0)$ belongs to the domain of the generator.

**Theorem 8.4** (Semigroup solves the ODE)**.** *Assume* $\{S(t)\}_{t\geq 0}$ *is a contraction semigroup, and* $u \in D(A)$*. Then for all* $t \geq 0$*:*

*(a)* $S(t)u \in D(A)$ *(solution flow preserves the domain)*

*(b)* $AS(t)u = S(t)Au$ *(solution operator commutes with the generator)*

*(c)* $\frac{d}{dt}S(t)u = AS(t)u$ *(semigroup solves* $\frac{dx}{dt} = Ax$ *with* $x(0) = u$*)*

*(d)* $t \mapsto S(t)u$ *is* $C^1$*-smooth (continuously differentiable solution curves)*

*Proof.* For parts (a) and (b), we observe that

$$\begin{aligned}
S(t)Au &= S(t)\lim_{s\to 0+}\frac{S(s)u - u}{s} && \text{by definition of } A \\
&= \lim_{s\to 0+}\frac{S(t)S(s)u - S(t)u}{s} && \text{by continuity (boundedness) of } S(t) \\
&= \lim_{s\to 0+}\frac{S(s)S(t)u - S(t)u}{s} && \text{by the semigroup property in Definition 8.1} \\
&= A\big(S(t)u\big) && \text{by definition of } A.
\end{aligned}$$

For part (c), fix $t > 0$ and first compute the right difference quotient:

$$\begin{aligned}
\lim_{h\to 0+}\frac{S(t+h)u - S(t)u}{h} &= \lim_{h\to 0+}\frac{S(h)S(t)u - S(t)u}{h} \\
&= A\big(S(t)u\big),
\end{aligned}$$

using from part (a) that $S(t)u$ belongs to the domain of $A$. Then compute the left difference quotient:

$$\begin{aligned}\lim_{h\to 0+} \frac{S(t-h)u - S(t)u}{-h} &= \lim_{h\to 0+} S(t-h)\Big(\frac{S(h)u-u}{h}\Big) \\ &= \lim_{h\to 0+} \Big\{S(t-h)\Big(\frac{S(h)u-u}{h} - Au\Big) + S(t-h)Au\Big\}.\end{aligned}$$

The first term is bounded in norm by

$$\|S(t-h)\| \left\| \frac{S(h)u-u}{h} - Au \right\| \to 0$$

as $h \to 0+$, using the contraction bound $\|S(\cdot)\| \leq 1$ and the definition of $A$. The second term converges as $h \to 0+$ to $S(t)Au = A\big(S(t)u\big)$.

Finally, for part (d) the solution curve is differentiable by part (c) and the derivative equals $S(t)\big(Au\big)$ by part (b). This derivative is a continuous function of $t$ by the continuity of solution paths in Definition 8.1. □

Next we show that the generator of a contraction semigroup is a closed operator having dense domain.

**Proposition 8.5.** *Assume* $\{S(t)\}_{t\geq 0}$ *is a contraction semigroup with generator* $A$. *Then the domain* $D(A)$ *is dense in* $X$, *and* $A : D(A) \to X$ *is a closed operator.*

*Proof.* To prove density of the domain, the idea is to consider $x \in X$ and show that for each $t > 0$, the integral $\int_0^t S(s)x\,ds$ belongs to the domain of $A$; then dividing by $t$ and letting $t \to 0$ shows that $x$ can be approximated arbitrarily well by points in $D(A)$. Details are in [Evans, Section 7.4].

To prove the generator is closed, we suppose $x_k \in D(A), x_k \to x \in X, Ax_k \to y \in X$. The task is to show that $x \in D(A)$ and $Ax = y$. By $C^1$-smoothness of the semigroup and the derivative formula in Theorem 8.4 we have

$$S(t)x_k - S(0)x_k = \int_0^t S(s)Ax_k\,ds,$$

and letting $k \to \infty$ proves

$$S(t)x - S(0)x = \int_0^t S(s)y\,ds.$$

Hence

$$\lim_{t\to 0+} \frac{S(t)x - x}{t} = y,$$

which proves that $x \in D(A)$ with $Ax = y$. □

## Resolvents

Our next step is to understand the resolvent operator of the generator.

**Definition 8.6.** Suppose $A$ is a linear operator from a subspace $D(A)$ into $X$. The **resolvent set** of $A$ is

$$\rho(A) = \{\lambda \in \mathbb{R} : \lambda I - A \text{ is a bijection of } D(A) \text{ onto } X\}.$$

When $\lambda$ belongs to the resolvent set we define the **resolvent operator**

$$R_\lambda : X \to D(A) \subset X$$

by

$$R_\lambda = (\lambda I - A)^{-1}.$$

Obviously $R_\lambda$ is linear since $A$ is linear.

As an example, suppose $A$ is an $N \times N$ real symmetric matrix acting on $X = \mathbb{R}^N$. Then $\lambda \in \rho(A)$ if and only if $\lambda$ is not an eigenvalue of $A$. That is, the resolvent set is the complement of the spectrum.

**Lemma 8.7** (Basic properties of the resolvent operator)**.** *Suppose $A$ is a linear operator from a subspace $D(A)$ into $X$. Let $\lambda, \mu \in \rho(A)$. Then the **resolvent identity** holds:*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu,$$

*and so resolvent operators commute:*

$$R_\lambda R_\mu = R_\mu R_\lambda.$$

*Further,*

$$A R_\lambda = \lambda R_\lambda - I. \tag{8.1}$$

*Proof.* For the resolvent identity, observe that $R_\mu^{-1} - R_\lambda^{-1} = (\mu - \lambda) I$ on $D(A)$ and then precompose with $R_\mu$ and postcompose with $R_\lambda$.

The resolvent identity implies commutativity of resolvents when $\lambda \neq \mu$, and of course there is nothing to prove when $\lambda = \mu$.

For formula (8.1), we note

$$\lambda R_\lambda - A R_\lambda = (\lambda I - A) R_\lambda = I.$$

□

**Lemma 8.8** (Continuity of the resolvent operator)**.** *Suppose* $A$ *is a closed linear operator from a subspace* $D(A)$ *into* $X$*. Fix* $\lambda \in \rho(A)$*.*

*Then* $R_\lambda : X \to X$ *is bounded and linear and commutes with* $A$*:*

$$R_\lambda A x = A R_\lambda x, \qquad x \in D(A).$$

*Proof.* We will show $R_\lambda$ is a closed operator. So suppose $x_k \in X, x_k \to x \in X, R_\lambda x_k \to y \in X$. The task is to prove that $R_\lambda x = y$. Let $y_k = R_\lambda x_k$, so that $y_k \in D(A), y_k \to y \in X$, and

$$\begin{aligned} A y_k &= A R_\lambda x_k \\ &= \lambda R_\lambda x_k - x_k \qquad \text{by (8.1)} \\ &= \lambda y_k - x_k \\ &\to \lambda y - x. \end{aligned}$$

Closedness of $A$ implies that $y \in D(A)$ with $Ay = \lambda y - x$. Hence $(\lambda I - A)y = x$, so that $R_\lambda x = y$.

Thus $R_\lambda$ is a closed operator, and so it is continuous (that is, bounded) by the Closed Graph Theorem from functional analysis.

For the commutativity, let $x \in D(A)$ and observe that

$$\begin{aligned} R_\lambda A x &= R_\lambda\big(\lambda I - (\lambda I - A)\big)x \\ &= \lambda R_\lambda x - x \\ &= A R_\lambda x \end{aligned}$$

by Lemma 8.7. □

**Proposition 8.9** (Resolvent of a contraction semigroup)**.** *If* $A$ *is the generator of a contraction semigroup then*

- $\rho(A) \supset (0, \infty)$,
- $R_\lambda x = \int_0^\infty e^{-\lambda t} S(t) x \, dt$ *for all* $x \in X, \lambda > 0$,
- $\|R_\lambda\| \leq 1/\lambda$ *for all* $\lambda > 0$.

Notice this resolvent formula is the Laplace transform of the semigroup.

*Proof.* See [Evans, Section 7.4]. The essence of the result is captured already by our ODE example, where $A$ is a real symmetric matrix with all eigenvalues less than or equal to $0$. If $\lambda > 0$ then all eigenvalues of $\lambda I - A$ are positive (as we see by diagonalizing). In particular, $\lambda I - A$ is invertible and so $\lambda \in \rho(A)$. Further, we compute for $u \in \mathbb{R}^N$ that

$$\begin{aligned}\int_0^\infty e^{-\lambda t} S(t) u \, dt &= \int_0^\infty e^{-(\lambda I - A)t} u \, dt \\ &= -(\lambda I - A)^{-1} e^{-(\lambda I - A)t} u \Big|_{t=0}^{t=\infty} \\ &= (\lambda I - A)^{-1} u \qquad \text{since } e^{-(\lambda I - A)t} u \to 0 \text{ as } t \to \infty \\ &= R_\lambda u.\end{aligned}$$

Hence

$$\|R_\lambda u\| \le \int_0^\infty e^{-\lambda t} \|S(t)u\| \, dt \le \frac{1}{\lambda} \|u\|$$

since $\|S(t)\| \le 1$. □

## 8.2 Characterization of contraction semigroups

Finally we arrive at the central result of the chapter.

**Theorem 8.10** (Hille–Yosida characterization of contraction semigroups)**.**
$A$ *is the generator of a contraction semigroup* $\{S(t)\}_{t \ge 0}$ *if and only if:*

- $A$ *is a closed linear operator,*
- *its domain* $D(A)$ *is a dense subspace of* $X$*,*
- *the resolvent set* $\rho(A)$ *contains all positive real numbers,*
- $\|R_\lambda\| \le 1/\lambda$ *for all* $\lambda > 0$.

*Proof.* The "necessary" direction $\Longrightarrow$ follows from Propositions 8.5 and 8.9.

The "sufficient" direction $\Longleftarrow$ is needed later for applications to PDEs, and so we give its proof in detail. The intuition is that we want to define the semigroup in terms of $A$ by letting $S(t) = e^{tA}$, but this exponential is not well defined, since $A$ is not necessarily a bounded operator. We work around this obstacle by first regularizing $A$.

Step 1 — Regularize the generator. Given $\lambda > 0$, we define a linear map $A_\lambda : X \to X$ by

$$A_\lambda = \lambda(\lambda R_\lambda - I),$$

so that $A_\lambda$ is bounded by hypothesis on the boundedness of $R_\lambda$. We have an alternative formula

$$A_\lambda = \lambda A R_\lambda = \lambda R_\lambda A,$$

by (8.1) and Lemma 8.8. This last expression explains why $A_\lambda$ should be regarded as a regularization of $A$, because formally it gives

$$A_\lambda = \frac{\lambda A}{\lambda I - A} \to A$$

as $\lambda \to \infty$.

To prove this regularization rigorously, we consider $x \in D(A)$ and observe by (8.1) and Lemma 8.8 that

$$\lambda R_\lambda x - x = AR_\lambda x = R_\lambda Ax \to 0 \qquad \text{as } \lambda \to \infty,$$

since $\|R_\lambda\| \leq 1/\lambda$. Hence

$$\lambda R_\lambda x \to x \qquad \text{as } \lambda \to \infty,$$

for each $x \in D(A)$. This convergence holds for each $x \in X$, since $D(A)$ is dense and $\|\lambda R_\lambda\| \leq 1$. Replacing $x$ with $Ax$ and recalling the definition of $A_\lambda$ therefore implies

$$A_\lambda x = \lambda R_\lambda(Ax) \to Ax \qquad \text{as } \lambda \to \infty,$$

for each $x \in D(A)$.

Step 2 — Semigroup for the regularized operator. Since $A_\lambda$ is bounded on $X$, we can define its exponential by a power series: let

$$\begin{aligned} S_\lambda(t) &= e^{tA_\lambda} \\ &= e^{-t\lambda I} e^{t\lambda^2 R_\lambda} \qquad \text{by the law of exponents, since the exponents commute,} \\ &= e^{-t\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda^2)^k R_\lambda^k}{k!} \end{aligned}$$

by the exponential series. One checks easily that $\{S_\lambda(t)\}_{t\geq 0}$ is a contraction semigroup: clearly $S_\lambda(0) = I$, and the semigroup property follows from the law of exponents, while the contraction property follows by estimating

$$\|S_\lambda(t)\| \leq e^{-t\lambda} \sum_{k=0}^{\infty} \frac{(t\lambda^2)^k}{k!} \left(\frac{1}{\lambda}\right)^k = e^{-t\lambda} e^{t\lambda} = 1.$$

The semigroup has generator $A_\lambda$, since for all $x \in X$ we have

$$\begin{aligned} \lim_{t\to 0+} \frac{S_\lambda(t)x - x}{t} &= \lim_{t\to 0+} \frac{e^{tA_\lambda}x - x}{t} \\ &= \lim_{t\to 0+} \sum_{k=1}^{\infty} \frac{t^{k-1} A_\lambda^k}{k!} x = A_\lambda x. \end{aligned}$$

Step 3 — Passing to the limit. Fix $T > 0$ and let $\lambda, \mu > 0$. Then $A_\lambda A_\mu = A_\mu A_\lambda$ by the Resolvent Identity, and so

$$A_\mu S_\lambda(s) = S_\lambda(s) A_\mu, \qquad s > 0. \tag{8.2}$$

We have for each $t \in [0, T]$ and $x \in X$ that

$$\begin{aligned} S_\lambda(t)x - S_\mu(t)x &= \int_0^t \frac{d}{ds}\Big(S_\mu(t-s)S_\lambda(s)x\Big)\,ds \\ &= \int_0^t \Big(S_\mu(t-s)S_\lambda(s)A_\lambda x - S_\mu(t-s)A_\mu S_\lambda(s)x\Big)\,ds \quad \text{by the product rule} \\ &= \int_0^t S_\mu(t-s)S_\lambda(s)(A_\lambda x - A_\mu x)\,ds \end{aligned}$$

by (8.2). Taking norms and using the contraction property, we find that $S_\lambda(t)x$ forms a Cauchy sequence with respect to $\lambda$:

$$\|S_\lambda(t)x - S_\mu(t)x\| \le T\|A_\lambda x - A_\mu x\| \to 0$$

as $\lambda, \mu \to \infty$ by the conclusion of Step 1, provided $x \in D(A)$. (Further, the convergence is uniform with respect to $t \in [0, T]$, for each fixed $x \in D(A)$.) Hence we may define

$$S(t)x = \lim_{\lambda\to\infty} S_\lambda(t)x, \qquad \forall\, t \in [0, T], \quad x \in D(A).$$

Since each $S_\lambda$ has norm at most $1$, the limit defining $S(t)x$ exists for all $x \in X$ (and the convergence is uniform with respect to $t \in [0, T]$, for each fixed $x \in X$). One now easily verifies that $\{S(t)\}_{t\ge0}$ satisfies the properties of a contraction semigroup.

Step 4 — Finding the generator. Write $B$ for the generator of $\{S(t)\}_{t\ge0}$. We will show $B = A$. For $x \in X$,

$$\begin{aligned} S_\lambda(t)x - x &= \int_0^t \Big(\frac{d}{ds}S_\lambda(s)x\Big)\,ds \\ &= \int_0^t S_\lambda(s)A_\lambda x\,ds. \end{aligned}$$

Letting $\lambda \to \infty$ and using the conclusion of Step 1 gives that

$$\frac{S(t)x - x}{t} = \frac{1}{t}\int_0^t S(s)Ax\,ds, \qquad x \in D(A),$$

where we have also divided by $t$. Letting $t \to 0+$, we deduce that the limit of the left side exists and equals the limit of the right side, which is $Ax$. Hence the domain of the generator $B$ contains $D(A)$, and $B = A$ there.

We must show the domain of $B$ equals $D(A)$. Let $\lambda > 0$. Then

$$\begin{aligned} D(B) &= (\lambda I - B)^{-1}(X) && \text{since } \lambda \in \rho(B) \text{ by Proposition 8.9} \\ &= (\lambda I - B)^{-1}\big((\lambda I - A)(D(A))\big) && \text{since } \lambda \in \rho(A) \text{ by hypothesis} \\ &= D(A) \end{aligned}$$

because $B = A$ on $D(A)$, by above. □

## 8.3 Dissipative operators

The resolvent conditions in Hille–Yosida's Theorem 8.10 can be simplified considerably. To avoid some technical complications, we assume in this section that $X$ is a Hilbert space.

**Definition 8.11.** A linear operator $A$ with domain $D(A)$ is **dissipative** if

$$\operatorname{Re}\langle x, Ax\rangle \le 0, \qquad x \in D(A),$$

and is **conservative** if equality holds for all $x \in D(A)$.

The point of this definition is that if $A$ is dissipative and

$$\frac{dx}{dt} = Ax, \quad x(0) = x_0,$$

then the norm-energy is dissipated:

$$\frac{d}{dt}\|x(t)\|^2 = 2\operatorname{Re}\big\langle x(t), x'(t)\big\rangle = 2\operatorname{Re}\langle x, Ax\rangle \le 0.$$

The solution $x(t)$ is unique when $A$ is dissipative, because if $\widetilde{x}$ is another solution with the same initial condition then the difference $y = x - \widetilde{x}$ satisfies $dy/dt = Ay$ and $y(0) = 0$, so that $\|y(t)\|^2 \le \|y(0)\|^2 = 0$, which implies $y(t) = 0$ for all $t \ge 0$.

*Example* 8.12. Consider

$$X = L^2(U), \qquad A = \Delta, \qquad D(A) = H^2 \cap H^1_0(U).$$

Then $A$ is dissipative because

$$\begin{aligned} \langle u, \Delta u\rangle_{L^2} &= -\int_U |\nabla u|^2\,dx && \text{by Green's Theorem, since } u = 0 \text{ on } \partial U, \\ &\le 0. \end{aligned}$$

This example corresponds to dissipation of $L^2$-energy for the heat equation: if $u_t = \Delta u$ then

$$\frac{d}{dt}\int_U u^2\,dx = 2\int_U u u_t\,dx = 2\int_U u\Delta u\,dx = -2\int_U |\nabla u|^2\,dx.$$

**Proposition 8.13** (Characterization of dissipation). *$A$ is dissipative if and only if*

$$\|(\lambda I - A)x\| \geq \lambda\|x\|, \qquad \forall x \in D(A), \quad \lambda > 0.$$

*Proof.* "$\Longleftarrow$":

$$\begin{aligned} 0 &\geq \lambda^2\|x\|^2 - \|(\lambda I - A)x\|^2 \\ &= 2\lambda \operatorname{Re}\langle x, Ax\rangle - \|Ax\|^2. \end{aligned}$$

Dividing by $\lambda$ and letting $\lambda \to \infty$ shows $\operatorname{Re}\langle x, Ax\rangle \leq 0$, so that $A$ is dissipative.
"$\Longrightarrow$": For all $x \in D(A)$ and $\lambda > 0$ we have

$$\|(\lambda I - A)x\|\|x\| \geq \operatorname{Re}\big\langle(\lambda I - A)x, x\big\rangle = \lambda\|x\|^2 - \operatorname{Re}\langle Ax, x\rangle \geq \lambda\|x\|^2,$$

where the final inequality holds because $A$ is dissipative. Dividing by $\|x\|$ now gives the desired condition.

□

Now we arrive at the promised simplification of the Hille–Yosida conditions, in a theorem due to Phillips.

**Theorem 8.14** (Contraction semigroups have dissipative generators). *Assume $A$ is a linear operator with dense domain $D(A)$ in the Hilbert space $X$.*

*(a) If $A$ is dissipative and $(\lambda_0 I - A)\big(D(A)\big) = X$ ("full range") for some $\lambda_0 > 0$, then $A$ is the generator of a contraction semigroup.*

*(b) If $A$ is the generator of a contraction semigroup, then $(\lambda I - A)\big(D(A)\big) = X$ for all $\lambda > 0$ and $A$ is dissipative*

Thus to obtain a contraction semigroup, one need only check the dissipation and full range conditions in part (a).

*Proof.* Part (a).

Step 1 — $A$ is closed. Proposition 8.13 implies that $\lambda_0 I - A$ is injective, and it is surjective by hypothesis. Hence $\lambda_0 I - A : D(A) \to X$ is a bijection, and its inverse map

$$(\lambda_0 I - A)^{-1} : X \to D(A) \subset X$$

is bounded by Proposition 8.13, so that the inverse map is continuous. In particular, the inverse map is closed, meaning its graph is a closed subset of $X \times X$. Therefore the original map $\lambda_0 I - A$ is closed (by a brief exercise), and thus $A$ is closed.

Step 2 — Full range. Denote by $\Lambda$ the set of $\lambda$-values for which $\lambda I - A$ has full range:

$$\Lambda = \{\lambda > 0 : (\lambda I - A)\big(D(A)\big) = X\}.$$

We want to show $\Lambda = (0, \infty)$.

First we show $\Lambda$ is open. Suppose $\lambda \in \Lambda$, so that $(\lambda I - A)^{-1} : X \to D(A) \subset X$ is bounded by the argument in Step 1. Suppose $\mu > 0$ is close enough to $\lambda$ that

$$|\lambda - \mu| \, \|(\lambda I - A)^{-1}\| < 1.$$

Then the mapping $I - (\lambda - \mu)(\lambda I - A)^{-1} : X \to X$ is a bijection, with inverse given by the geometric series (or *Neumann series*)

$$\big(I - (\lambda - \mu)(\lambda I - A)^{-1}\big)^{-1} = \sum_{k=0}^{\infty} (\lambda - \mu)^k (\lambda I - A)^{-k}.$$

In particular the mapping is a surjection: $\big(I - (\lambda - \mu)(\lambda I - A)^{-1}\big)(X) = X$. Hence $\mu \in \Lambda$, because

$$\begin{aligned}
(\mu I - A)\big(D(A)\big) &= \big((\lambda I - A) - (\lambda - \mu)I\big)\big(D(A)\big) \\
&= \big(I - (\lambda - \mu)(\lambda I - A)^{-1}\big)(\lambda I - A)\big(D(A)\big) \\
&= \big(I - (\lambda - \mu)(\lambda I - A)^{-1}\big)(X) \qquad \text{since } \lambda \in \Lambda \\
&= X.
\end{aligned}$$

Thus we have shown $\mu \in \Lambda$ for all $\mu$ sufficiently close to $\lambda$, and so $\Lambda$ is an open set.

Next we show $\Lambda$ is a closed subset of $(0, \infty)$. Suppose $\lambda_n \in \Lambda$ with $\lambda_n \to \lambda > 0$. To prove $\lambda \in \Lambda$, we must show $\lambda I - A$ has full range.

Consider $y \in X$. Since $\lambda_n \in \Lambda$ we can find $x_n \in D(A)$ satisfying

$$(\lambda_n I - A)x_n = y. \tag{8.3}$$

We now show how to pass to the limit in this equation. Observe that the sequence $\{x_n\}$ is bounded since

$$\|x_n\| \leq \frac{1}{\lambda_n} \|(\lambda_n I - A)x_n\| = \frac{1}{\lambda_n}\|y\|$$

by Proposition 8.13. To prove $\{x_n\}$ is a Cauchy sequence we estimate

$$\begin{aligned}
\lambda_m \|x_n - x_m\| &\leq \|\lambda_m I(x_n - x_m) - A(x_n - x_m)\| && \text{by Proposition 8.13} \\
&= \|\lambda_m x_n - A x_n - y\| && \text{by (8.3) with "}m\text{"} \\
&= |\lambda_m - \lambda_n| \, \|x_n\| && \text{by (8.3) with "}n\text{"} \\
&\to 0 \qquad \text{as } m, n \to \infty.
\end{aligned}$$

Since $\{x_n\}$ is Cauchy, the sequence converges to some $x \in X$, and then from (8.3) we find $Ax_n \to \lambda x - y$. Since $A$ is closed (by Step 1), we deduce that $x \in D(A)$ and $Ax = \lambda x - y$, so that

$$(\lambda I - A)x = y.$$

Thus $\lambda I - A$ has full range, and so $\lambda \in \Lambda$ as we wanted to show.

The set $\Lambda$ is known to be open, closed, and nonempty (containing $\lambda_0$), and so it equals the whole interval $(0, \infty)$.

Step 3 — Apply Hille-Yosida. Recall $D(A)$ is dense by hypothesis, and is closed by Step 1. Each $\lambda > 0$ belongs to the resolvent set of $A$, since $\lambda I - A$ is surjective by Step 2 and injective by Proposition 8.13. That Proposition gives also the resolvent norm bound $\|R_\lambda\| \leq 1/\lambda$. Thus the Hille–Yosida Theorem 8.10 implies that $A$ generates a contraction semigroup.

Part (b). Fix $\lambda > 0$. The Hille–Yosida Theorem 8.10 insures that $\lambda \in \rho(A)$, and so $(\lambda I - A)\big(D(A)\big) = X$.

To show $A$ is dissipative, note that for all $x \in D(A)$ we have

$$\operatorname{Re} \langle S(t)x, x \rangle \leq \|S(t)x\| \|x\| \leq \|x\|^2$$

and hence for all $t > 0$,

$$\operatorname{Re} \left\langle \frac{S(t)x - x}{t}, x \right\rangle \leq 0.$$

Letting $t \to 0$ shows $\operatorname{Re} \langle Ax, x \rangle \leq 0$, so that $A$ is dissipative. □

***Exercise*** 8.3. Show that if $A$ is dissipative and $\lambda_0 > 0$ is such that

$$(\lambda_0 I - A)\big(D(A)\big) = X$$

(the "full range" condition) then $D(A)$ is dense in $X$. Thus when applying Phillips' Theorem 8.14(a), one need not verify the density of $D(A)$.

## 8.4 *Application:* Solving parabolic, hyperbolic and Schrödinger PDEs by semigroups

Take $U$ to be a bounded domain in $\mathbb{R}^N$.

***Exercise*** 8.4. Consider the homogeneous **parabolic** problem

$$\begin{aligned} u_t + Lu &= 0 \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0, T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}, && \text{IC} \end{aligned}$$

where $L$ is a time-independent elliptic operator in divergence form, with no lower order terms:

$$Lu = -\sum_{j,k} \left(a^{jk}(x)u_{x_j}\right)_{x_k}.$$

Use semigroup methods (Theorem 8.14) to obtain an existence and uniqueness result for this problem. Then state your result carefully, including appropriate hypotheses on $a^{jk}, U, g$.

*Hint.* Let $X = L^2(U)$ and $A = -L$, and choose a function space $D(A)$ on which $A$ makes sense. When you attempt to prove $A$ is dissipative, you might find that you must further restrict $D(A)$.

***Exercise*** 8.5. Suppose $\{S(t)\}_{t\geq 0}$ is a contraction semigroup, so that $x(t) = S(t)u$ is a $C^1$-solution of

$$\frac{dx}{dt} = Ax$$

for $t \geq 0$, with initial condition $x(0) = u \in D(A)$.

Now fix $\beta \in \mathbb{R}$ and find a $C^1$-solution of

$$\frac{dy}{dt} = Ay + \beta y$$

for $t \geq 0$, with initial condition $y(0) = u \in D(A)$. *Hint.* Two-line proof.

***Exercise*** 8.6. Repeat Exercise 8.4 for the operator $L$ that includes a 0-th order term:

$$Lu = -\sum_{j,k} \left(a^{jk}(x)u_{x_j}\right)_{x_k} + c(x)u.$$

Proceed in two steps:

(i) First assume $c$ is bounded and nonnegative.

(ii) Then assume $c$ is bounded but could change sign.

***Exercise*** 8.7. Consider the homogeneous **hyperbolic** problem

$$\begin{aligned} u_{tt} + Lu &= 0 \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0,T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}, && \text{IC} \\ u_t &= h \quad \text{on } U \times \{0\}, && \text{IC} \end{aligned}$$

where $L$ is a time-independent elliptic operator in divergence form, with no lower order terms:

$$Lu = -\sum_{j,k} \left(a^{jk}(x)u_{x_j}\right)_{x_k}.$$

(i) Show that the second order PDE can be rewritten as a system of two first order PDEs, namely as

$$w_t = Aw$$

where

$$w = \begin{pmatrix} u \\ u_t \end{pmatrix}, \qquad A = \begin{pmatrix} 0 & 1 \\ -L & 0 \end{pmatrix}.$$

(ii) Define

$$D(A) = \left[H^2 \cap H_0^1(U)\right] \times H_0^1(U), \qquad X = H_0^1(U) \times L^2(U).$$

Then $X$ is a Hilbert space with inner product

$$\langle w, \widetilde{w}\rangle_X = \int_U \Big( \sum_{j,k} a^{jk} u_{x_j} \widetilde{u}_{x_k} + v\widetilde{v}\Big)\, dx,$$

where we use the notation

$$w = \begin{pmatrix} u \\ v \end{pmatrix}, \qquad \widetilde{w} = \begin{pmatrix} \widetilde{u} \\ \widetilde{v} \end{pmatrix}.$$

Prove $A$ is dissipative. (In fact it is *conservative*.)

(iii) Show that $I - A$ has full range, that is, show it maps $D(A)$ onto $X$.

(iv) Use semigroup methods (Theorem 8.14) to deduce an existence and uniqueness result for the hyperbolic problem. State your result carefully, including appropriate hypotheses on $a^{jk}, U, g, h$.

*Example* 8.15. Consider the homogeneous **Schrödinger** problem

$$\begin{aligned} iu_t + Lu &= 0 \quad \text{in } U_T, && \text{PDE} \\ u &= 0 \quad \text{on } \partial U \times (0,T], && \text{BC} \\ u &= g \quad \text{on } U \times \{0\}, && \text{IC} \end{aligned}$$

where $L$ is a time-independent symmetric elliptic operator in divergence form, with zeroth order term:

$$Lu = -\sum_{j,k} \left(a^{jk}(x) u_{x_j}\right)_{x_k} + V(x)u.$$

Here $u$ is a complex valued "wave function", with $|u|^2$ representing the probability density for the location of a quantum particle in the domain $U$. Previously we have written $c(x)$ for the zeroth order term, but in quantum mechanics this term represents an external electric potential, and so we write it as $V(x)$. We assume $V(x)$ is bounded and real valued, and that $a^{jk} \in C^1(\overline{U})$.

The Schrödinger equation can be rewritten as

$$u_t = iLu$$

and so we let $A = iL$ with domain $D(A) = H^2 \cap H_0^1(U;\mathbb{C})$, and $X = L^2(U;\mathbb{C})$, where the "$\mathbb{C}$" reminds us that our functions take complex values. As usual, $D(A)$ is dense in $X$ because both spaces contain the smooth functions with compact support.We show $A$ is conservative, as follows. For $u \in D(A)$, repeated integration by parts (using that $u = 0$ on the boundary and $V$ is real-valued) implies that $L$ is "formally selfadjoint", with

$$\langle Lu, u\rangle = \langle u, Lu\rangle = \overline{\langle Lu, u\rangle}, \qquad u \in D(A).$$

(Here we use the complex inner product, $\langle f, g\rangle = \int_U f\overline{g}\,dx$.) Thus $\langle Lu, u\rangle$ is real, and so

$$\mathrm{Re}\,\langle Au, u\rangle = \mathrm{Re}\ i\langle Lu, u\rangle = 0,$$

which means $A$ is conservative. Remember the quantity being conserved is the $L^2$ norm of $u$, which we interpret in this case as conservation of the total probability:

$$\frac{d}{dt}\|u\|_{L^2}^2 = 2\,\mathrm{Re}\,\langle u, \frac{du}{dt}\rangle_{L^2} = 2\,\mathrm{Re}\langle u, Au\rangle_{L^2} = 0.$$

Next we show $I - A$ has full range. Denote by $\{u_j\}$ an ONB of eigenfunctions for $L$, with eigenvalues $\{\lambda_j\}$. To solve the equation

$$(I - A)u = f$$

where $f \in L^2(U)$ is arbitrary, we substitute $u = \sum_j c_j u_j$ and $f = \sum_j \langle f, u_j\rangle_{L^2} u_j$, and find that

$$c_j = \frac{\langle f, u_j\rangle_{L^2}}{1 - i\lambda_j}.$$

The denominator is nonzero, since the eigenvalues are real. The series $u = \sum_j c_j u_j$ converges in $L^2(U;\mathbb{C})$ and in $H_0^1(U;\mathbb{C})$, by arguing with the synthesis operator as in the proof of Theorem 5.1. Hence one can verify that $u$ satisfies $(I - A)u = f$ weakly. Elliptic regularity theory further implies that $u \in H^2(U;\mathbb{C})$ and hence that $(I-A)u = f$ classically. (In order to invoke elliptic regularity, we assume that $U$ has $C^1$-smooth boundary.) Thus $I - A$ has full range.

Existence of a contraction semigroup that solves the Schrödinger equation now follows from Theorem 8.14.

*Remark* 8.16. For symmetric operators $L$ with an ONB of eigenfunctions, the time evolution problems treated so far in this chapter can be handled more easily by eigenfunction expansions, rather than by semigroups. For example, to solve the parabolic equation $u_t + Lu = 0$ with Dirichlet boundary conditions and initial condition $u = g$, one may simply separate variables with

$$u(x,t) = \sum_j e^{-\lambda_j t} c_j u_j(x)$$

where $c_j = \langle g, u_j \rangle_{L^2}$. That is, $u = S(t)g$ where the semigroup operator (or "solution operator") for the parabolic problem is given in terms of the eigenfunctions and eigenvalues of $L$ by

$$S(t)g = \sum_j e^{-\lambda_j t} \langle g, u_j \rangle_{L^2} \, u_j.$$

For the Schrödinger problem, the corresponding formula is

$$S(t)g = \sum_j e^{i\lambda_j t} \langle g, u_j \rangle_{L^2} \, u_j.$$

Of course, in this separation of variables approach one would still need to justify that these series converge and give weak solutions of the PDE.

So why use semigroups? Three reasons:

1. Semigroups provide a unifying approach and a shorthand notation for solving linear evolutionary problems.

2. Semigroups immediately solve the parabolic/hyperbolic problem as soon as we know the elliptic problem is solvable, regardless of whether or not the elliptic operator has an ONB of eigenfunctions.

3. Semigroups provide a conceptual framework for solving **nonhomogeneous** and **nonlinear** evolutionary problems, as we illustrate in the next section.

## 8.5 *Application:* Nonhomogeneous and nonlinear evolution equations

Until now, our semigroup methods have applied only to homogeneous evolution equations, such as a diffusion equation with no sources or a wave equation with zero forcing. Now we explain how to handle nonhomogeneous equations too, using semigroups.

**Theorem 8.17** (Nonhomogeneous linear ODE). *Assume* $A$ *is a linear operator with dense domain* $D(A)$ *in the Hilbert space* $X$. *Suppose* $A$ *is dissipative and* $(\lambda_0 I - A)\big(D(A)\big) = X$ *for some* $\lambda_0 > 0$. *Let* $f \in C^1([0,T];X)$ *and* $g \in D(A)$.

*Then the nonhomogeneous ODE*

$$\frac{dx}{dt} = Ax + f, \qquad x(0) = g,$$

*has a unique* $C^1$*-solution for* $t \in [0,T]$. *The solution is given by the Duhamel formula*

$$x(t) = S(t)g + \int_0^t S(t-s)f(s)\,ds,$$

*where* $\{S(t)\}_{t \geq 0}$ *is the semigroup generated by* $A$.

The Duhamel formula expresses the solution as a response to the initial condition plus the sum (integral) of responses to the forcing at each time $s \in [0,t]$. You have probably seen a simple version of the formula already in an ODE course, where $x(t)$ is real valued, $A \in \mathbb{R}$ is constant, and $S(t) = e^{At}$: if you solve the resulting first order ODE by the method of integrating factors, then you will obtain Duhamel's formula in this special case.

*Proof.* Obviously $x(0) = g$. To prove the ODE, use the semigroup properties, product rule, and fundamental theorem of calculus to show formally that

$$\begin{aligned} \frac{dx}{dt} &= AS(t)g + \int_0^t AS(t-s)f(s)\,ds + S(t-t)f(t) \\ &= Ax + f \end{aligned}$$

since $S(0) = I$. These calculations can be justified rigorously by a short argument with difference quotients. □

To illustrate the Duhamel formula, we finish the course with a nonlinear diffusion equation, which we solve with the help of semigroups and the contraction mapping principle from Chapter 1.

*Example* 8.18 (Nonlinear diffusion equation). Suppose $F : \mathbb{R} \to \mathbb{R}$ is a Lipschitz continuous function with Lipschitz constant $K$. Consider the nonlinear diffusion problem

$$\begin{aligned} u_t - \Delta u &= F(u) && \text{in } U_T, && \text{PDE} \\ u &= 0 && \text{on } \partial U \times (0,T], && \text{BC} \\ u &= g && \text{on } U \times \{0\}, && \text{IC} \end{aligned}$$

and write $S(t)$ for the semigroup associated with $A = \Delta$ for $D(A) = H^2 \cap H^1_0(U), X = L^2(U)$. We want to solve

$$\frac{du}{dt} = Au + F(u)$$

with $u(0) = g$. Assume a solution exists, regard the nonlinear term $F(u)$ in the PDE as a nonhomogeneity, and apply the Duhamel formula:

$$u(t) = S(t)g + \int_0^t S(t-s)F(u(s))\,ds, \qquad t \in [0,T].$$

This formula does not provide us with a solution, or course, since the unknown function $u \in C\big([0,T]; L^2(U)\big)$ appears on both the left and right sides of the equation. Denote the right side of the equation by $(Zu)(t)$. One can check that $Zu \in C\big([0,T]; L^2(U)\big)$, meaning $t \mapsto (Zu)(t)$ is a continuous function taking values in $L^2(U)$. Thus the Duhamel formula says that we seek a fixed point of the operator $Z$:

$$u = Zu,$$

where we regard $Z$ as acting on the Banach space $C\big([0,T]; L^2(U)\big)$ equipped with the max-norm with respect to $t$.

We will show $Z$ is a (strict) contraction, so that we may apply Banach's Fixed Point Theorem 1.4. To prove $Z$ is a contraction we adapt the proof of Picard's Theorem 1.6, proving a short-time contraction estimate as follows:

$$\begin{aligned}
\max_{t\in[0,T]} \|Zu_1(t) - Zu_2(t)\|_{L^2} &\le \max_{t\in[0,T]} \int_0^t \|S(t-s)\| \|F(u_1(s)) - F(u_2(s))\|_{L^2}\,ds \\
&\le \int_0^T K\|u_1(s) - u_2(s)\|_{L^2}\,ds \qquad \text{since } F \text{ is Lipschitz} \\
&\le TK \max_{s\in[0,T]} \|u_1(s) - u_2(s)\|_{L^2}.
\end{aligned}$$

The contraction constant $TK$ is less than $1$ provided the terminal time $T$ is small enough. Thus we have proved short-time existence of a solution for the nonlinear diffusion equation.

**Concluding remark.** With this final example we have brought the course full circle, and hinted at future developments. The vast and ever expanding world of nonlinear PDEs can be explored by combining techniques from hard analysis such as Sobolev spaces with tools from functional analysis such as semigroups and contraction mappings. *Bon voyage!*

***Exercise*** 8.8. Write a two page summary of the most important and memorable results and general techniques from this course. Be brief, but thoughtful — explain how these main results fit together (*e.g.* what did we use Sobolev inequalities for?).

You need not state results precisely or technically — intuitive explanations are more helpful at this stage.

# Bibliography

[BlanchardBrüning] P. Blanchard and E. Brüning. Variational methods in mathematical physics. A unified approach. Translated from the German by Gillian M. Hayes. Texts and Monographs in Physics. Springer–Verlag, Berlin, 1992.

[Brezis] H. Brezis. Functional analysis, Sobolev spaces and partial differential equations. Universitext. Springer, New York, 2011.

[Chicone] C. Chicone. Ordinary differential equations with applications. Second edition. Texts in Applied Mathematics, 34. Springer, New York, 2006.

[Evans] L. C. Evans. Partial Differential Equations. Second edition. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence, RI, 2010.

[Folland] G. B. Folland. Real analysis. Modern techniques and their applications. Second edition. Pure and Applied Mathematics (New York). A Wiley–Interscience Publication. John Wiley & Sons, Inc., New York, 1999.

[Keener] J. P. Keener. Principles of applied mathematics. Transformation and approximation. Revised edition. Perseus Books, Advanced Book Program, Cambridge, MA, 2000.

[Pazy] A. Pazy. Semigroups of linear operators and applications to partial differential equations. Applied Mathematical Sciences, 44. Springer–Verlag, New York, 1983.

[Zeidler] E. Zeidler. Nonlinear functional analysis and its applications. I. Fixed-point theorems. Translated from the German by Peter R. Wadsack. Springer–Verlag, New York, 1986.