Methods of Applied Mathematics

Todd Arbogast and Jerry Bona

Department of Mathematics

The University of Texas at Austin

Contents

CHAPTER 1

Preliminaries

1.1. Elementary topology

In applied mathematics, we are often faced with analyzing mathematical structures as they might relate to real-world phenomena. In applying mathematics, real phenomena or objects are conceptualized as abstract mathematical objects. Collections of such objects are called *sets*. The objects in a set of interest may also be related to each other; that is, there is some *structure* on the set. We call such structured sets *spaces*.

Examples. (1) A vector space (algebraic structure).

(2) The set of integers $\mathbb{Z}$ (number theoretical structure or arithmetic structure).

(3) The set of real numbers $\mathbb{R}$ or the set of complex numbers $\mathbb{C}$ (algebraic and topological structure).

We start the discussion of spaces by putting forward sets of "points" on which we can talk about the notions of *convergence* or *limits* and associated *continuity* of functions.

A simple example is a set X with a notion of distance between any two points of X. A sequence $\{x_n\}_{n=1}^{\infty} \subset X$ converges to $x \in X$ if the distance from x_n to x tends to 0 as n increases. This definition relies on the following formal concept.

Definition. A *metric* or *distance function* on a set is a function $d : X \times X \to \mathbb{R}$ satisfying:

1. (positivity) for any $x, y \in X$, $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$;
2. (symmetry) for any $x, y \in X$, $d(x, y) = d(y, x)$;
3. (triangle inequality) for any $x, y, z \in X$, $d(x, y) \leq d(x, z) + d(z, y)$.

A *metric space* (X, d) is a set X together with an associated metric $d : X \times X \to \mathbb{R}$.

Example. $(\mathbb{R}^d, |\cdot|)$ is a metric space, where for $x, y \in \mathbb{R}^d$, the distance from x to y is

$$|x - y| = \left\{ \sum_{i=1}^{d} (x_i - y_i)^2 \right\}^{1/2} .$$

It turns out that the notion of distance or metric is sometimes stronger than what actually appears in practice. The more fundamental concept upon which much of the mathematics developed here rests, is that of limits. That is, there are important spaces arising in applied mathematics that have well defined notions of limits, but these limiting processes are *not* compatible with any metric. We shall see such examples later; let it suffice for now to motivate a weaker definition of limits.

A sequence of points $\{x_n\}_{n=1}^{\infty}$ can be thought of as converging to x if every "neighborhood" of x contains all but finitely many of the x_n, where a neighborhood is a subset of points containing x that we think of as "close" to x. Such a structure is called a *topology*. It is formalized as follows.

Definition. A *topological space* $(X, \mathcal{T})$ is a nonempty set X of points with a family $\mathcal{T}$ of subsets, called *open*, with the properties:

(1) $X \in \mathcal{T}$, $\emptyset \in \mathcal{T}$;
(2) If $\omega_1, \omega_2 \in \mathcal{T}$, then $\omega_1 \cap \omega_2 \in \mathcal{T}$;
(3) If $\omega_\alpha \in \mathcal{T}$ for all α in some index set $\mathcal{I}$, then $\bigcup_{\alpha \in \mathcal{I}} \omega_\alpha \in \mathcal{T}$.

The family $\mathcal{T}$ is called a *topology* for X. Given $A \subset X$, we say that A is *closed* if its complement A^c is open.

Example. If X is any nonempty set, we can always define the two topologies:

(1) $\mathcal{T}_1 = \{\emptyset, X\}$, called the *trivial* topology;
(2) $\mathcal{T}_2$ consisting of the collection of all subsets of X, called the *discrete* topology.

Proposition 1.1. *The sets $\emptyset$ and X are both open and closed. Any finite intersection of open sets is open. Any intersection of closed sets is closed. The union of any finite number of closed sets is closed.*

Proof. We need only show the last two statements, as the first two follow directly from the definitions. Let $A_\alpha \subset X$ be closed for $\alpha \in \mathcal{I}$. Then one of deMorgan's laws gives that

$$\Big(\bigcap_\alpha A_\alpha\Big)^c = \bigcup_\alpha A_\alpha^c \quad \text{is open.}$$

Finally, if $\mathcal{J} \subset \mathcal{I}$ is finite, then

$$\Big(\bigcup_{\alpha \in \mathcal{J}} A_\alpha\Big)^c = \bigcap_{\alpha \in \mathcal{J}} A_\alpha^c \quad \text{is open.}$$

□

It is often convenient to define a simpler collection of open sets that immediately generates a topology.

Definition. Given a topological space $(X, \mathcal{T})$ and an $x \in X$, a *base for the topology at x* is a collection $\mathcal{B}_X$ of open sets containing x such that for any open $E \ni x$, there is $B \subset \mathcal{B}_X$ such that

$$x \in B \subset E\ .$$

A *base for the topology* $\mathcal{B}$ is a collection of open sets that contains a base at x for all $x \in X$.

Proposition 1.2. *A collection $\mathcal{B}$ of subsets of X is a base for a topology $\mathcal{T}$ if and only if* (1) *each $x \in X$ is contained in some $B \in \mathcal{B}$ and* (2) *if $x \in B_1 \cap B_2$ for $B_1, B_2 \in \mathcal{B}$, then there is some $B_3 \in \mathcal{B}$ such that $x \in B_3 \subset B_1 \cap B_2$. If* (1) *and* (2) *are valid, then*

$$\mathcal{T} = \{E \subset X : E \textit{ is a union of subsets in } \mathcal{B}\}\ .$$

Proof. ($\Rightarrow$) Since X and $B_1 \cap B_2$ are open, (1) and (2) follow from the definition of a base at x.

($\Leftarrow$) Let $\mathcal{T}$ be defined as above. Then $\emptyset \in \mathcal{T}$ (the vacuous union), $X \in \mathcal{T}$ by (1), and arbitrary unions of sets in $\mathcal{T}$ are again in $\mathcal{T}$. It remains to show the intersection property. Let $E_1, E_2 \in \mathcal{T}$, and $x \in E_1 \cap E_2$ (if $E_1 \cap E_2 = \emptyset$, there is nothing to prove). Then there are sets $B_1, B_2 \in \mathcal{B}$ such that

$$x \in B_1 \subset E_1\ , \qquad x \in B_2 \subset E_2\ ,$$

so

$$x \in B_1 \cap B_2 \subset E_1 \cap E_2 \ .$$

Now (2) gives $B_3 \in \mathcal{B}$ such that

$$x \in B_3 \subset E_1 \cap E_2 \ .$$

Thus $E_1 \cap E_2$ is a union of elements in $\mathcal{B}$, and is thus in $\mathcal{T}$. □

We remark that instead of using open sets, one can consider *neighborhoods* of points $x \in X$, which are sets $N \ni x$ such that there is an open set E satisfying $x \in E \subset N$.

THEOREM 1.3. *If (X, d) is a metric space, then $(X, \mathcal{T})$ is a topological space, where a base for the topology is given by*

$$\mathcal{T}_B = \{B_r(x) : x \in X \quad \text{and} \quad r > 0\} \ ,$$

where

$$B_r(x) = \{y \in X : d(x, y) < r\}$$

is the ball of radius r about x.

PROOF. Point (1) is clear. For (2), suppose $x \in B_r(y) \cap B_s(z)$. Then $x \in B_\rho(x) \subset B_r(y) \cap B_s(z)$, where $\rho = \frac{1}{2}\min(r - d(x, y), s - d(x, z)) > 0$. □

Thus metric spaces have a natural topological structure. However, not all topological spaces are induced as above by a metric, so the class of topological spaces is genuinely richer.

DEFINITION. Let $(X, \mathcal{T})$ be a topological space. The *closure* of $A \subset X$, denoted $\overline{A}$, is the intersection of all closed sets containing A:

$$\overline{A} = \bigcap_{\substack{F \text{ closed} \\ F \supseteq A}} F \ .$$

PROPOSITION 1.4. *$\overline{A}$ is closed, and is the smallest closed set containing A.*

PROOF. This follows by Proposition 1.1 and the definition. □

DEFINITION. The *interior* of $A \subset X$, denoted A°, is the union of all open sets contained in A:

$$A^\circ = \bigcup_{\substack{E \text{ open} \\ E \subset A}} E \ .$$

PROPOSITION 1.5. *A° is open, and is the largest open set contained in A.*

PROOF. This also follows from Proposition 1.1 and the definition. □

PROPOSITION 1.6. *$A \subset \bar{A}$, $\bar{\bar{A}} = \bar{A}$, $\overline{A \cup B} = \bar{A} \cup \bar{B}$, and A closed $\Leftrightarrow A = \bar{A}$.*
$A \supseteq A^\circ$, $A^{\circ\circ} = A^\circ$, $(A \cap B)^\circ = A^\circ \cap B^\circ$, and A open $\Leftrightarrow A = A^\circ$.

PROPOSITION 1.7. *$(A^c)^\circ = (\,\bar{A}\,)^c$, $(A^\circ)^c = \overline{(A^c)}$.*

Proof.

$$x \notin (\bar{A})^c \Leftrightarrow x \in \bar{A} \Leftrightarrow x \in \bigcap_{\substack{F \text{ closed} \\ F \supset A}} F \Leftrightarrow x \notin \Bigg(\bigcap_{\substack{F \text{ closed} \\ F \supset A}} F \Bigg)^c \Leftrightarrow x \notin \bigcup_{\substack{F^c \text{ open} \\ F^c \subset A^c}} F^c = (A^c)^\circ \ .$$

The second result is similar. □

Definition. A point $x \in X$ is an *accumulation point* of $A \subset X$ if every open set containing x intersects $A \setminus \{x\}$. Also, a point $x \in A$ is an *interior point* of A if there is some open set E such that

$$x \in E \subset A \ .$$

Finally, $x \in A$ is an *isolated point* if there is an open set $E \ni x$ such that $E \setminus \{x\} \cap A = \emptyset$.

Proposition 1.8. *For $A \subset X$, $\bar{A}$ is the union of the set of accumulation points of A and A itself and A^0 is the union of the interior points of A.*

Proof. Exercise. □

Definition. A set $A \subset X$ is *dense* in X if $\bar{A} = X$.

Definition. The *boundary* of $A \subset X$, denoted ∂A, is

$$\partial A = \bar{A} \cap \overline{A^c} \ .$$

Proposition 1.9. *If $A \subset X$, then ∂A is closed and*

$$\bar{A} = A^\circ \cup \partial A \ , \qquad A^\circ \cap \partial A = \emptyset \ .$$

Moreover,

$$\partial A = \partial A^c = \{x \in X : \text{ every open } E \ni x \text{ intersects both } A \text{ and } A^c\} \ .$$

Proof. Exercise. □

Definition. A sequence $\{x_n\}_{n=1}^\infty \subset X$ *converges* to $x \in X$, or has *limit* x, if given any open $E \ni x$, there is $N > 0$ such that $x_n \in E$ for all $n \geq N$ (i.e., the entire tail of the sequence is contained in E).

Proposition 1.10. *If $\lim_{n\to\infty} x_n = x$, then x is an accumulation point of $\{x_n\}_{n=1}^\infty$, interpreted as a set.*

Proof. Exercise. □

We remark that if x is an accumulation point of $\{x_n\}_{n=1}^\infty$, there may be no subsequence $\{x_{n_k}\}_{k=1}^\infty$ converging to x.

Example. Let X be the set of nonnegative integers, and a base $\mathcal{T}_B = \{\{0, 1, \dots, i\}$ for each $i \geq 1\}$. Then $\{x_n\}_{n=1}^\infty$ with $x_n = n$ has 0 as an accumulation point, but no subsequence converges to 0.

If $x_n \to x \in X$ and $x_n \to y \in X$, it is possible that $x \neq y$.

Example. Let $X = \{a, b\}$ and $\mathcal{T} = \{\emptyset, \{a\}, \{a, b\}\}$. Then the sequence $x_n = a$ for all n converges to both a and b.

Definition. A topological space $(X, \mathcal{T})$ is called *Hausdorff* if given distinct $x, y \in X$, there are disjoint open sets E_1 and E_2 such that $x \in E_1$ and $y \in E_2$.

PROPOSITION 1.11. *If $(X, \mathcal{T})$ is Hausdorff, then every set consisting of a single point is closed. Moreover, limits of sequences are unique.*

PROOF. Exercise. $\square$

DEFINITION. A point $x \in X$ is a *strict limit point* of $A \subset X$ if there is a sequence $\{x_n\} \subset A \setminus \{x\}$ such that $\lim_{n\to\infty} x_n = x$.

PROPOSITION 1.12. *Every $x \in \partial A$ is either an isolated point, or a strict limit point of A and A^c.*

PROOF. Exercise. $\square$

Note that if x is an isolated point of X, then $x \notin A^\circ$, so $\partial A \neq \partial A^\circ$ in general.

Metric spaces are less suseptible to pathology than general topological spaces.

PROPOSITION 1.13. *If (X, d) is a metric space and $\{x_n\}_{n=1}^\infty$ is a sequence in X, then $x_n \to x$ if and only if, given $\varepsilon > 0$, there is $N > 0$ such that*

$$d(x, x_n) < \varepsilon \qquad \forall\ n \geq N\ .$$

That is, $x_n \in B_\varepsilon(x)$ for all $n \geq N$.

PROOF. If $x_n \to x$, then the tail of the sequence is in every open set $E \ni x$. In particular, this holds for the open sets $B_\varepsilon(x)$. Conversely, if E is any open set containing x, then the open balls at x form a base for the topology, so there is some $B_\varepsilon(x) \subset E$ which contains the tail of the sequence. $\square$

PROPOSITION 1.14. *Every metric space is Hausdorff.*

PROOF. Exercise. $\square$

PROPOSITION 1.15. *If (x, d) is a metric space and $A \subset X$ has an accumulation point x, Then there is some sequence $\{x_n\}_{n=1}^\infty \subset A$ such that $x_n \to x$.*

PROOF. Given integer $n \geq 1$, there is some $x_n \in B_{1/n}(x)$, since x is an accumulation point. Thus $x_n \to x$. $\square$

We avoid problems arising with limits in general topological spaces by the following definition of continuity.

DEFINITION. A mapping f of a topological space $(X, \mathcal{T})$ into a topological space $(Y, \mathcal{S})$ is *continuous* if the inverse image of every open set in Y is open in X.

This agrees with our notion of continuity on $\mathbb{R}$.

We say that f is continuous at a point $x \in X$ if given any open set $E \subset Y$ containing $f(x)$, then $f^{-1}(E)$ contains an open set D containing x. That is,

$$x \in D \quad \text{and} \quad f(D) \subset E\ .$$

A map is continuous if and only if it is continuous at each point of X.

PROPOSITION 1.16. *If $f : X \to Y$ and $g : Y \to Z$ are continuous, then $g \circ f : X \to Z$ is continuous.*

PROOF. Exercise. $\square$

PROPOSITION 1.17. *If f is continuous and $x_n \to x$, then $f(x_n) \to f(x)$.*

Proof. Exercise. $\square$

The converse of Proposition 1.17 is false in general. When the hypothesis $x_n \to x$ always implies $f(x_n) \to f(x)$, we say that f is *sequentially continuous*.

Proposition 1.18. *If $f : X \to Y$ is sequentially continuous, and if X is a metric space, then f is continuous.*

Proof. Let $E \subset Y$ be open and $A = f^{-1}(E)$. We must show that A is open. Suppose not. Then there is some $x \in A$ such that $B_r(x) \not\subset A$ for all $r > 0$. Thus for $r_n = 1/n$, $n \geq 1$ an integer, there is some $x_n \in B_{r_n}(x) \cap A^c$. Since $x_n \to x$, $f(x_n) \to f(x) \in E$. But $f(x_n) \in E^c$ for all n, so $f(x)$ is an accumulation point of E^c. That is, $f(x) \in \overline{E^c} \cap \overline{E} = \partial E$. Hence, $f(x) \in \partial E \cap E = \partial E \cap E^\circ = \emptyset$, a contradiction. $\square$

Suppose we have a map $f : X \to Y$ that is both injective (one to one) and surjective (onto), such that both f and f^{-1} are continuous. Then f and f^{-1} map open sets to open sets. That is $E \subset X$ is open if and only if $f(E) \subset Y$ is open. Therefore $f(\mathcal{T}) = \mathcal{S}$, and, from a topological point of view, X and Y are indistinguishable. Any topological property of X is shared by Y, and conversely. For example, if $x_n \to x$ in X, then $f(x_n) \to f(x)$ in Y, and conversely ($y_n \to y$ in $Y \Rightarrow f^{-1}(y_n) \to f^{-1}(y)$ in X).

Definition. A *homeomorphism* between two topological spaces X and Y is a one-to-one continuous mapping f of X onto Y for which f^{-1} is also continuous. If there is a homeomorphism $f : X \to Y$, we say that X and Y are *homeomorphic*.

It is possible to define two or more nonhomeomorphic topologies on any set X of at least two points. If $(X, \mathcal{T})$ and $(X, \mathcal{S})$ are topological spaces, and $\mathcal{S} \supset \mathcal{T}$, then we say that $\mathcal{S}$ is *stronger* than $\mathcal{T}$ or that $\mathcal{T}$ is *weaker* than $\mathcal{S}$.

Example. The trivial topology is weaker than any other topology. The discrete topology is stronger than any other topology.

Proposition 1.19. *The topology $\mathcal{S}$ is stronger than $\mathcal{T}$ if and only if the identity mapping $I : (X, \mathcal{S}) \to (X, \mathcal{T})$ is continuous.*

Proposition 1.20. *Given a collection $\mathcal{C}$ of subsets of X, there is a weakest topology $\mathcal{T}$ containing $\mathcal{C}$.*

Proof. Since the intersection of topologies is again a topology (prove this),

$$\mathcal{C} \subset \mathcal{T} = \bigcap_{\substack{\mathcal{S} \supseteq \mathcal{C} \\ \mathcal{S} \text{ a topology}}} \mathcal{S}$$

is the weakest such topology (which is nonempty since the discrete topology is a topology containing $\mathcal{C}$). $\square$

Given a topological space $(X, \mathcal{T})$ and $A \subset X$, we obtain a topology $\mathcal{S}$ on A by restriction. We say that this topology on A is *inherited* from X. Specifically

$$\mathcal{S} = \mathcal{T} \cap A \equiv \{E \subset A : \text{ there is some } G \subset \mathcal{T} \text{ such that } E = A \cap G\} .$$

That $\mathcal{S}$ is a topology on A is easily verified. We also say that A is a *subspace* of X.

Given two topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{S})$, we can define a topology $\mathcal{R}$ on

$$X \times Y = \{(x, y) : x \in X, \ y \in Y\} ,$$

called the *product topology*, from the base

$$\mathcal{R}_B = \{E_1 \times E_2 : E_1 \in \mathcal{T},\ E_2 \in \mathcal{S}\}\ .$$

It is easily verified that this is indeed a base; moreover, we could replace $\mathcal{T}$ and $\mathcal{S}$ by bases and obtain the same topology $\mathcal{R}$.

EXAMPLE. If (X, d_1) and (X, d_2) are metric spaces, then a base for $X \times Y$ is

$$\{B_r(x) \times B_s(y) : x \in X,\ y \in Y, \ \text{ and } \ r, s > 0\}\ .$$

Moreover, $d : (X \times Y) \times (X \times Y) \to \mathbb{R}$ defined by

$$d((x_1, y_1), (x_2, y_2)) = d_1(x_1, x_2) + d_2(y_1, y_2)$$

is a metric that gives the same topology.

EXAMPLE. $\mathbb{R}^2$ has two equivalent and natural bases for the usual Euclidean topology, the set of all (open) circles, and the set of all (open) rectangles.

This construction can be generalized to obtain an arbitrary product of spaces. Let $(X_\alpha, \mathcal{T}_\alpha)$ $\alpha \in \mathcal{I}$ be a collection of topological spaces. Then $X = \times_{\alpha\in\mathcal{I}} X_\alpha$, defined to be the collection of all points $\{x_\alpha\}_{\alpha\in\mathcal{I}}$ with the property that $x_\alpha \in X_\alpha$ for all $\alpha \in \mathcal{I}$, has a *product topology* with base

$$\mathcal{T}_B = \Big\{ \mathop{\times}_{\alpha\in\mathcal{I}} E_\alpha : E_\alpha \in \mathcal{T}_\alpha \ \forall\ \alpha \in \mathcal{I} \ \text{ and } \ E_\alpha = X_\alpha \\ \text{for all but a finite number of } \alpha \in \mathcal{I} \Big\}\ .$$

The *projection* map $\pi_\alpha : X \to X_\alpha$ is defined for $x = \{x_\beta\}_{\beta\in\mathcal{I}}$ by $\pi_\alpha x = x_\alpha$, which gives the α-th coordinate of x.

REMARK. The notation $\{x_\alpha\}_{\alpha\in\mathcal{I}}$ is properly understood as a map $g : \mathcal{I} \to \cup_{\alpha\in\mathcal{I}} X_\alpha$, where $g(\alpha) = x_\alpha \in X_\alpha$ for all $\alpha \in \mathcal{I}$. Then $X = \times_{\alpha\in\mathcal{I}} X_\alpha$ is the collection of all such maps, and $\pi_\alpha(g) = g(\alpha)$ is evaluation at $\alpha \in \mathcal{I}$. However, we will continue to use the more informal view of X as consisting of "points" $\{x_\alpha\}_{\alpha\in\mathcal{I}}$.

PROPOSITION 1.21. *Each π_α is continuous. Furthermore, the product topology is the weakest topology on X that makes each π_α continuous.*

PROOF. If $E_\alpha \subset X_\alpha$ is open, then

$$\pi_\alpha^{-1}(E_\alpha) = \mathop{\times}_{\beta\in\mathcal{I}} E_\beta\ ,$$

where $E_\beta = X_\beta$ for $\beta \neq \alpha$, is a basic open set and so is open. Finite intersections of these sets must be open, and indeed these form our base. It is therefore obvious that the product topology as defined must form the weakest topology for which each π_α is continuous. □

PROPOSITION 1.22. *If X_α, $\alpha \in \mathcal{I}$, and Y are topological spaces, then a function $f : \times_{\alpha\in\mathcal{I}} X_\alpha \to Y$ is continuous if and only if $\pi_\alpha \circ f$ is continuous for each $\alpha \in \mathcal{I}$.*

PROOF. Exercise. □

PROPOSITION 1.23. *If X is Hausdorff and $A \subset X$, then A is Hausdorff (in the inherited topology). If $\{X_\alpha\}_{\alpha\in\mathcal{I}}$ are Hausdorff, then $\times_{\alpha\in\mathcal{I}} X_\alpha$ is Hausdorff (in the product topology).*

PROOF. Exercise. □

Most topologies of interest have an infinite number of open sets. For such spaces, it is often difficult to draw conclusions. However, there is an important class of topological space with a finiteness property.

DEFINITION. Let $(X,\mathcal{T})$ be a topological space and $A \subset X$. A collection $\{E_\alpha\}_{\alpha\in\mathcal{I}} \subset X$ is called an *open cover* of A if $A \subset \bigcup_{\alpha\in\mathcal{I}} E_\alpha$. If every open cover of A contains a finite subcover (i.e., the collection $\{E_\alpha\}$ can be reduced to a finite number of open sets that still cover A), then A is called *compact*.

An interesting point arises right away: Does the compactness of A depend upon the way it is a subset of X? Another way to ask this is, if $A \subsetneqq X$ is compact, is A compact when it is viewed as a subset of itself? That is, $(A, \mathcal{T} \cap A)$ is a topological space, and $A \subset A$, so is A also compact in this context? What about the converse? If A is compact in itself, is A compact in X? It is easy to verify that both these questions are answered in the affirmative. Thus compactness is a property of a set, independent of some larger space in which it may live.

The Heine-Borel Theorem states that every closed and bounded subset of $\mathbb{R}^d$ is compact, and conversely. The proof is technical and can be found in most introductory books on real analysis (such as the one by Royden [**Roy**] or Rudin [**Ru0**]).

PROPOSITION 1.24. *A closed subset of a compact space is compact. A compact subset of a Hausdorff space is closed.*

PROOF. Let X be compact, and $F \subset X$ closed. If $\{E_\alpha\}_{\alpha\in\mathcal{I}}$ is an open cover of F, then $\{E_\alpha\}_{\alpha\in\mathcal{I}} \cup F^c$ is an open cover of X. By compactness, there is a finite subcover $\{E_\alpha\}_{\alpha\in\mathcal{J}} \cup F^c$. But then $\{E_\alpha\}_{\alpha\in\mathcal{J}}$ covers F, so F is compact.

Suppose X is Hausdorff and $K \subset X$ is compact. (We write $K \subset\subset X$ in this case, and read it as "K compactly contained in X.") We claim that K^c is open. Fix $y \in K^c$. For each $x \in K$, there are open sets E_x and G_x such that $x \in E_x$, $y \in G_x$, and $E_x \cap G_x = \emptyset$, since X is Hausdorff. The sets $\{E_x\}_{x\in K}$ form an open cover of K, so a finite subcollection $\{E_x\}_{x\in A}$ still covers K. Thus

$$E = \bigcap_{x\in A} G_x$$

is open, contains y, and does not intersect K. Since y is arbitrary, K^c is open and therefore K closed. □

PROPOSITION 1.25. *The continuous image of a compact set is compact.*

PROOF. Exercise. □

An amazing fact about compact spaces is contained in the following theorem. Its proof can be found in most introductory texts in analysis or topology (see [**Roy**], [**Ru1**]).

THEOREM 1.26 (Tychonoff). *Let $\{X_\alpha\}_{\alpha\in\mathcal{I}}$ be an indexed family of compact topological spaces. Then the product space $X = \times_{\alpha\in\mathcal{I}} X_\alpha$ is compact in the product topology.*

A common way to use compactness in metric spaces is contained in the following result.

PROPOSITION 1.27. *If X is a compact metric space and $\{x_n\}_{n=1}^\infty$ is a sequence in X, then there is a subsequence $\{x_{n_k}\}_{k=1}^\infty$ which converges in X.*

PROOF. Suppose not. Then the sets

$$F_n = \{x_n, x_{n+1}, x_{n+2}, \dots\}$$

have no limit points, which are accumulation points in a metric space, so F_n is closed. Now

$$\bigcap_{n=1}^{\infty} F_n = \emptyset \ ,$$

so $\{F_n^c\}_{n=1}^{\infty}$ forms an open cover of X. Hence, for some N, $\{F_n^c\}_{n=1}^{N}$ covers X. But

$$x_N \notin \bigcup_{n=1}^{N} F_n^c = F_N^c \ .$$

This contradiction establishes the result. □

1.2. Lebesgue measure and integration

The Riemann integral is quite satisfactory for continuous functions, or functions with not too many discontinuities, defined on bounded subsets of $\mathbb{R}^d$; however, it is not so satisfactory for discontinuous functions, nor can it be easily generalized to functions defined on sets outside $\mathbb{R}^d$, such as probability spaces. Measure theory resolves these difficulties. It seeks to measure the size of relatively arbitrary subsets of some set X. From such a well defined notion of size, the integral can be defined. We summarize the basic theory here, but omit most of the proofs. They can be found in most texts in real analysis (see e.g., [**Roy**], [**Ru0**], [**Ru2**]).

It turns out that a consistent measure of subset size cannot be defined for all subsets of a set X. We must either modify our notion of size or restrict to only certain types of subsets. The latter course appears a good one since, as we will see, the subsets of $\mathbb{R}^d$ that can be measured include any set that can be approximated well via rectangles.

Definition. A collection $\mathcal{A}$ of subsets of X is called a *σ-algebra* on X if

i) $X \in \mathcal{A}$;
ii) whenever $A \in \mathcal{A}$, $A^c \in \mathcal{A}$;
iii) whenever $A_n \in \mathcal{A}$ for $n = 1, 2, 3, \ldots$ (i.e., countably many A_n), then also $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Proposition 1.28.

i) *$\emptyset \in \mathcal{A}$.*
ii) *If $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots$, then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.*
iii) *If $A, B \in \mathcal{A}$, then $A \setminus B = A \cap B^c \in \mathcal{A}$.*

Proof. Exercise. □

Definition. By a *measure* on $\mathcal{A}$, we mean a function $\mu : \mathcal{A} \to R$, where $R = [0, +\infty]$ for a *positive measure* with $\mu \not\equiv +\infty$ and $R = \mathbb{C}$ for a *complex measure*, which is *countably additive*. This means that if $A_n \in \mathcal{A}$ for $n = 1, 2, \ldots$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, then

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) \ .$$

That is, the size or *measure* of a set is the sum of the measures of countably many disjoint pieces of the set that fill it up.

Proposition 1.29.

i) *$\mu(\emptyset) = 0$.*

ii) *If* $A_n \in \mathcal{A}$, $n = 1, 2, \ldots, N$ *are pairwise disjoint, then*

$$\mu\biggl(\bigcup_{n=1}^{N} A_n\biggr) = \sum_{n=1}^{N} \mu(A_n) \ .$$

iii) *If* μ *is a positive measure and* $A, B \in \mathcal{A}$ *with* $A \subset B$, *then*

$$\mu(A) \le \mu(B) \ .$$

iv) *If* $A_n \in \mathcal{A}$, $n = 1, 2, \ldots,$ *and* $A_n \subset A_{n+1}$ *for all* n, *then*

$$\mu\biggl(\bigcup_{n=1}^{\infty} A_n\biggr) = \lim_{n\to\infty} \mu(A_n) \ .$$

v) *If* $A_n \in \mathcal{A}$, $n = 1, 2, \ldots$, $\mu(A_1) < \infty$, *and* $A_n \supseteq A_{n+1}$ *for all* n, *then*

$$\mu\biggl(\bigcap_{n=1}^{\infty} A_n\biggr) = \lim_{n\to\infty} \mu(A_n) \ .$$

Proof. i) Since $\mu \not\equiv +\infty$, there is $A \in \mathcal{A}$ such that $\mu(A)$ is finite. Now $A = A \cup \bigcup_{i=1}^{\infty} \emptyset$, so $\mu(A) = \mu(A) + \sum_{i=1}^{\infty} \mu(\emptyset)$. Thus $\mu(\emptyset) = 0$.

ii) Let $A_n = \emptyset$ for $n > N$. Then

$$\mu\biggl(\bigcup_{n=1}^{N} A_n\biggr) = \mu\biggl(\bigcup_{n=1}^{\infty} A_n\biggr) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{N} \mu(A_n) \ .$$

iii) Let $C = B \setminus A$. Then $C \cap A = \emptyset$, so

$$\mu(A) + \mu(C) = \mu(C \cup A) = \mu(B) \ ,$$

and $\mu(C) \ge 0$ gives the result.

iv) Let $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \ge 2$. Then the $\{B_n\}$ are pairwise disjoint, and, for any $N \le \infty$,

$$A_N = \bigcup_{n=1}^{N} A_n = \bigcup_{n=1}^{N} B_n \ ,$$

so

$$\begin{aligned}
\mu\biggl(\bigcup_{n=1}^{\infty} A_n\biggr) &= \mu\biggl(\bigcup_{n=1}^{\infty} B_n\biggr) = \sum_{n=1}^{\infty} \mu(B_n) \\
&= \lim_{N\to\infty} \sum_{n=1}^{N} \mu(B_n) = \lim_{N\to\infty} \mu\biggl(\bigcup_{n=1}^{N} B_n\biggr) \\
&= \lim_{N\to\infty} \mu(A_N) \ .
\end{aligned}$$

v) Let $B_n = A_n \setminus A_{n+1}$ and $B = \bigcap_{n=1}^{\infty} A_n$. Then the B_n and B are pairwise disjoint,

$$A_N = A_1 \setminus \bigcup_{n=1}^{N-1} B_n \ , \qquad \text{and} \qquad A_1 = B \cup \bigcup_{n=1}^{\infty} B_n \ .$$

In consequence of the countable additivity,

$$\mu(A_1) = \mu(B) + \sum_{n=1}^{\infty} \mu(B_n) < \infty \ ,$$

or

$$\begin{aligned}\mu(B) &= \mu(A_1) - \sum_{n=1}^{N-1} \mu(B_n) - \sum_{n=N}^{\infty} \mu(B_n) \\ &= \mu(A_N) - \sum_{n=N}^{\infty} \mu(B_n) \ .\end{aligned}$$

Since the series $\sum_{n=1}^{\infty} \mu(B_n)$ converges, the limit as $N \to \infty$ of the second term on the right-hand side of the last equation is zero and the result follows. $\square$

A triple consisting of a set X, a σ-algebra $\mathcal{A}$ of subsets of X, and a measure μ defined on $\mathcal{A}$, i.e., $(X, \mathcal{A}, \mu)$, is called a *measure space*.

An important σ-algebra is one generated by a topology, namely the family $\mathcal{B}$ of all Borel sets in $\mathbb{R}^d$.

DEFINITION. The *Borel sets* $\mathcal{B}$ in $\mathbb{R}^d$ is the smallest family of subsets of $\mathbb{R}^d$ with the properties:

i) each open set is in $\mathcal{B}$;
ii) if $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$;
iii) if $\{A_n\}_{n=1}^{\infty} \subset \mathcal{B}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$.

That is, $\mathcal{B}$ contains all open sets and is closed under complements and countable unions.

That there is such a smallest family follows from the facts that the family of all subsets satisfies (ii)–(iii), and if $\{\mathcal{B}_\alpha\}_{\alpha \in \mathcal{I}}$ is any collection of families satisfying (i)–(iii), then $\bigcap_{\alpha \in \mathcal{I}} \mathcal{B}_\alpha$ also satisfies (i)–(iii).

Note that closed sets are in $\mathcal{B}$, as well as countable intersections by deMorgan's rule. Obviously, $\mathcal{B}$ is a σ-algebra.

REMARK. This definition makes sense relative to the open sets in any topological space.

THEOREM 1.30. *There exists a unique positive measure μ, called* Lebesgue measure*, defined on the Borel sets $\mathcal{B}$ of $\mathbb{R}^d$, having the properties that if $A \subset \mathcal{B}$ is a* rectangle*, i.e., there are numbers a_i and b_i such that*

$$A = \{x \in \mathbb{R}^d : a_i < x_i \text{ or } a_i \le x_i \text{ and } x_i < b_i \text{ or } x_i \le b_i \ \forall\ i\} \ ,$$

then $\mu(A) = \prod_{i=1}^{d}(b_i - a_i)$ and μ is translation invariant, which means that if $x \in \mathbb{R}^d$ and $A \in \mathcal{B}$, then

$$\mu(x + A) = \mu(A) \ ,$$

where $x + A = \{y \in \mathbb{R}^d : y = x + z \text{ for some } z \in A\} \in \mathcal{B}$.

The construction of Lebesgue measure is somewhat tedious, and can be found in most texts in real analysis (see, e.g., [**Roy**], [**Ru0**], [**Ru2**]). Note that an interesting point arising in this theorem is to determine why $x + A \in \mathcal{B}$ if $A \in \mathcal{B}$. This follows since the mapping $f(y) = y + x$ is a homeomorphism of $\mathbb{R}^d$ onto $\mathbb{R}^d$, and hence preserves the open sets which generate the Borel sets.

A dilemma arises. If $A \in \mathcal{B}$ is such that $\mu(A) = 0$, we say A is a *set of measure zero*. As an example, a $(d-1)$-dimensional hyperplane has d-dimensional measure zero. If we intersect the hyperplane with $A \subset \mathbb{R}^d$, the measure should be zero; however, such an intersection may not be a Borel set. We would like to say that if $\mu(A) = 0$ and $B \subset A$, then μ applies to B and $\mu(B) = 0$.

Let the sets of measure zero be

$$\mathcal{Z} = \{A \subset \mathbb{R}^d : \exists\, B \in \mathcal{B} \ \text{ with } \ \mu(B) = 0 \ \text{ and } \ A \subset B\} ,$$

and define the *Lebesgue measurable sets* $\mathcal{M}$ to be

$$\mathcal{M} = \{A \subset \mathbb{R}^d : \exists\, B \in \mathcal{B},\ Z_1, Z_2 \in \mathcal{Z} \ \text{ such that } \ A = (B \cup Z_1) \setminus Z_2\} .$$

We leave it to the reader to verify that $\mathcal{M}$ is a σ-algebra.

Next extend $\mu : \mathcal{M} \to [0, \infty]$ by

$$\mu(A) = \mu(B)$$

where $A = (B \cup Z_1) \setminus Z_2$ for some $B \in \mathcal{B}$ and $Z_1, Z_2 \in \mathcal{Z}$. That this definition is independent of the decomposition is easily verified, since $\mu|_{\mathcal{Z}} = 0$.

Thus we have

Theorem 1.31. *There exists a σ-algebra $\mathcal{M}$ of subsets of $\mathbb{R}^d$ and a positive measure $\mu : \mathcal{M} \to [0, \infty]$ satisfying the following.*

i) *Every open set in $\mathbb{R}^d$ is in $\mathcal{M}$.*
ii) *If $A \subset B \in \mathcal{M}$ and $\mu(B) = 0$, then $A \in \mathcal{M}$ and $\mu(A) = 0$.*
iii) *If A is a rectangle with x_i bounded between a_i and b_i, then $\mu(A) = \prod_{i=1}^d (b_i - a_i)$.*
iv) *μ is translation invariant: if $x \in \mathbb{R}^d$, $A \in \mathcal{M}$, then $x + A \in \mathcal{M}$ and $\mu(A) = \mu(x + A)$.*

Sets outside $\mathcal{M}$ exist, and are called *unmeasurable* or *non-measurable* sets. We shall not meet any in this course. Moreover, for practical purposes, we might simply restrict $\mathcal{M}$ to $\mathcal{B}$ in the following theory with only minor technical differences.

We now consider functions defined on measure spaces, taking values in the extended real number system $\overline{\mathbb{R}} \equiv \mathbb{R} \cup \{-\infty, +\infty\}$, or in $\mathbb{C}$.

Definition. Suppose $\Omega \subset \mathbb{R}^d$ is measurable. A function $f : \Omega \to \overline{\mathbb{R}}$ is *measurable* if the inverse image of every open set in $\mathbb{R}$ is measurable. A function $g : \Omega \to \mathbb{C}$ is measurable if its real and imaginary parts are measurable.

We remark that measurability depends on $\mathcal{M}$, but not on μ! It would be enough to verify that the sets

$$E_\alpha = \{x \in \Omega : f(x) > \alpha\}$$

are measurable for all $\alpha \in \mathbb{R}$ to conclude that f is measurable.

Theorem 1.32.

i) *If f and g are measurable, so are $f + g$, $f - g$, fg, $\max(f, g)$, and $\min(f, g)$.*
ii) *If f is measurable and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $g \circ f$ is measurable.*
iii) *If f is defined on $\Omega \subset \mathbb{R}^d$, f continuous, and Ω measurable, then f is measurable.*
iv) *If $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions, then*

$$\inf_n f_n \ , \quad \sup_n f_n \ , \quad \liminf_{n\to\infty} f_n \ , \quad \textit{and} \quad \limsup_{n\to\infty} f_n$$

are measurable functions.

COROLLARY 1.33. *If f is measurable, then so are*

$$f^+ = \max(f, 0) \ , \quad f^- = -\min(f, 0) \ , \quad \text{and} \ \ |f| \ .$$

Moreover, if $\{f_n\}_{n=1}^\infty$ are measurable and converge pointwise, the limit function is measurable.

REMARK. With these definitions, $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

DEFINITION. If X is a set and $E \subset X$, then the function $\mathcal{X}_E : X \to \mathbb{R}$ given by

$$\mathcal{X}_E(x) = \begin{cases} 1 & \text{if } x \in E \ , \\ 0 & \text{if } x \notin E \ , \end{cases}$$

is called the *characteristic function* of E. If $s : X \to \mathbb{R}$ has finite range, then s is called a *simple* function.

Of course, if the range of s is $\{c_1, \dots, c_n\}$ and

$$E_i = \{x \in X : s(x) = c_i\} \ ,$$

then

$$s(x) = \sum_{i=1}^n c_i \mathcal{X}_{E_i}(x) \ ,$$

and s is measurable if and only if each E_i is measurable.

Every function can be approximated by simple functions.

THEOREM 1.34. *Given any function $f : \Omega \subset \mathbb{R}^d \to \overline{\mathbb{R}}$, there is a sequence $\{s_n\}_{n=1}^\infty$ of simple functions such that*

$$\lim_{n\to\infty} s_n(x) = f(x) \quad \text{for any} \ \ x \in \Omega$$

(i.e., s_n converges pointwise to f). If f is measurable, the $\{s_n\}$ can be chosen measurable. Moreover, if f is bounded, $\{s_n\}$ can be chosen so that the convergence is uniform. If $f \geq 0$, then the $\{s_n\}$ may be chosen to be monotonically increasing at each point.

PROOF. If $f \geq 0$, define for $n = 1, 2, \dots$ and $i = 1, 2, \dots, n2^n$,

$$E_{n,i} = \left\{ x \in \Omega : \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n} \right\} \ ,$$
$$F_n = \{x \in \Omega : f(x) \geq n\} \ .$$

Then

$$s_n(x) = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathcal{X}_{E_{n,i}}(x) + n\mathcal{X}_{F_n}$$

has the desired properties. In the general case, let $f = f^+ - f^-$ and approximate f^+ and f^- as above. □

It is now straightforward to define the Lebesgue integral. Let $\Omega \subset \mathbb{R}^d$ be measurable and $s : \Omega \to \mathbb{R}$ be a measurable simple function given as

$$s(x) = \sum_{i=1}^n c_i \mathcal{X}_{E_i}(x) \ .$$

Then we define the Lebesgue integral of s over Ω to be

$$\int_\Omega s(x)\,dx = \sum_{i=1}^n c_i\mu(E_i)\ .$$

If $f : \Omega \to [0,\infty]$ is measurable, we define

$$\int_\Omega f(x)\,dx = \sup_s \int_\Omega s(x)\,dx\ ,$$

where the supremum is taken over all measurable functions satisfying $0 \le s(x) \le f(x)$ for $x \in \Omega$. Note that the integral of f may be $+\infty$.

If f is measurable and real-valued, then $f = f^+ - f^-$, where $f^+ \ge 0$ and $f^- \ge 0$. In this case, define

$$\int_\Omega f(x)\,dx = \int_\Omega f^+(x)\,dx - \int_\Omega f^-(x)\,dx\ ,$$

provided at least one of the two integrals on the right is finite.

Finally, if f is complex-valued, apply the above construction to the real and imaginary parts of f, provided the integrals of these parts are finite.

Definition. We say that a real-valued measurable function f is *integrable* if the integrals of f^+ and f^- are both finite. If only one is finite, then f is not integrable; however, in that case we assign $+\infty$ or $-\infty$ to the integral.

Proposition 1.35. *The real-valued measurable function f is* integrable *over Ω if and only if*

$$\int_\Omega |f(x)|\,dx < \infty\ .$$

Definition. The class of all integrable functions on $\Omega \subset \mathbb{R}^d$, Ω measurable, is denoted

$$\mathcal{L}(\Omega) = \{\text{measurable } f : \int_\Omega |f(x)|\,dx < \infty\}\ .$$

Theorem 1.36. *If f is Riemann integrable on a compact set $K \subset \mathbb{R}^d$, then $f \in \mathcal{L}(K)$ and the Riemann and Lebesgue integrals agree.*

Certain properties of the Lebesgue integral are clear from its definition.

Proposition 1.37. *Assume that all functions and sets appearing below are measurable.*

(a) *If $|f|$ is bounded on Ω and $\mu(\Omega) < \infty$, then $f \in \mathcal{L}(\Omega)$.*

(b) *If $a \le f \le b$ on Ω and $\mu(\Omega) < \infty$, then*

$$a\mu(\Omega) \le \int_\Omega f(x)\,dx \le b\mu(\Omega)\ .$$

(c) *If $f \le g$ on Ω, then*

$$\int_\Omega f(x)\,dx \le \int_\Omega g(x)\,dx\ .$$

(d) *If $f, g \in \mathcal{L}(\Omega)$, then $f + g \in \mathcal{L}(\Omega)$ and*

$$\int_\Omega (f+g)(x)\,dx = \int_\Omega f(x)\,dx + \int_\Omega g(x)\,dx\ .$$

(e) *If* $f \in \mathcal{L}(\Omega)$ *and* $c \in \mathbb{R}$ *(or* $\mathbb{C}$*), then*

$$\int_\Omega cf(x)\,dx = c\int_\Omega f(x)\,dx\ .$$

(f) *If* $f \in \mathcal{L}(\Omega)$*, then* $|f| \in \mathcal{L}(\Omega)$ *and*

$$\Big|\int_\Omega f(x)\,dx\Big| \le \int_\Omega |f(x)|\,dx\ .$$

(g) *If* $f \in \mathcal{L}(\Omega)$ *and* $A \subset \Omega$*, then* $f \in \mathcal{L}(A)$*. If also* $f \ge 0$*, then*

$$0 \le \int_A f(x)\,dx \le \int_\Omega f(x)\,dx\ .$$

(h) *If* $\mu(\Omega) = 0$*, then*

$$\int_\Omega f(x)\,dx = 0\ .$$

(i) *If* $f \in \mathcal{L}(\Omega)$ *and* $\Omega = A \cup B$*,* $A \cap B = \emptyset$*, then*

$$\int_\Omega f(x)\,dx = \int_A f(x)\,dx + \int_B f(x)\,dx\ .$$

Part (i) has a natural and useful generalization.

THEOREM 1.38. *If* $f \in \mathcal{L}(\Omega)$*,* $A \subset \Omega$*,* $A_n \in \mathcal{M}$ *for* $n = 1, 2, \dots$*,* $A_i \cap A_j = \emptyset$ *for* $i \ne j$*, and* $A = \bigcup_{n=1}^\infty A_n$*, then*

$$\int_A f(x)\,dx = \sum_{n=1}^\infty \int_{A_n} f(x)\,dx\ . \tag{1.1}$$

Moreover, if $f \ge 0$*, the function* $\lambda : \mathcal{M} \to \mathbb{R}$ *given by*

$$\lambda(A) = \int_A f(x)\,dx$$

is a positive measure.

PROOF. That λ is a positive measure follows from (1.1), which gives the countable additivity. If (1.1) is valid when $f \ge 0$, it will follow for any real or complex valued function via the decomposition $f = f_1 + if_2 = f_1^+ - f_1^- + i(f_2^+ - f_2^-)$, where $f_i^\pm \ge 0$.

For a characteristic function $\mathcal{X}_E$, E measurable, (1.1) holds since μ is countably additive:

$$\int_A \mathcal{X}_E(x)\,dx = \mu(A \cap E) = \sum_{n=1}^\infty \mu(A_n \cap E) = \sum_{n=1}^\infty \int_{A_n} \mathcal{X}_E(x)\,dx\ .$$

Because of (d) and (e) in Proposition 1.37, (1.1) also holds for any simple function.

If $f \ge 0$ and s is a simple function such that $0 \le s \le f$, then

$$\int_A s(x)\,dx = \sum_{n=1}^\infty \int_{A_n} s(x)\,dx \le \sum_{n=1}^\infty \int_{A_n} f(x)\,dx\ .$$

Thus

$$\int_A f(x)\,dx = \sup_{s\le f}\int_A s(x)\,dx \le \sum_{n=1}^\infty \int_{A_n} f(x)\,dx\ .$$

However, by iterating Proposition 1.37(i), it follows that

$$\sum_{k=1}^{n} \int_{A_k} f(x)\,dx = \int_{\bigcup_{k=1}^{n} A_k} f(x)\,dx \le \int_A f(x)\,dx$$

for any n. The last two inequalities imply (1.1) for f. $\square$

From Proposition 1.37(h),(i), it is clear that if A and B are measurable sets and $\mu(A \setminus B) = \mu(B \setminus A) = 0$, then

$$\int_A f(x)\,dx = \int_B f(x)\,dx$$

for any integrable f. Moreover, if f and g are integrable and $f(x) = g(x)$ for all $x \in A \setminus C$ where $\mu(C) = 0$, then

$$\int_A f(x)\,dx = \int_A g(x)\,dx \ .$$

Thus sets of measure zero are negligible in integration.

If a property P holds for every $x \in E \setminus A$ where $\mu(A) = 0$, then we say that P holds for *almost every* $x \in E$, or that P holds *almost everywhere* on E. We generally abbreviate "almost everywhere" as "a.e." (or "p.p." in French).

PROPOSITION 1.39. *If $f \in \mathcal{L}(\Omega)$, where Ω is measurable, and if*

$$\int_A f(x)\,dx = 0$$

for every measurable $A \subset \Omega$, then $f = 0$ a.e. on Ω.

PROOF. Suppose not. Decompose f as $f = f_1 + if_2 = f_1^+ - f_1^- + i(f_2^+ - f_2^-)$. At least one of $f_1^\pm$, $f_2^\pm$ is not zero a.e. Let g denote one such component of f. Thus $g \ge 0$ and g is not zero a.e. on Ω. However, $\int_A g(x)\,dx = 0$ for every measurable $A \subset \Omega$. Let

$$A_n = \{x \in \Omega : g(x) > 1/n\} \ .$$

Then $\mu(A_n) = 0 \ \forall\ n$ and $A_0 = \bigcup_{n=1}^\infty A_n = \{x \in \Omega : g(x) > 0\}$. But $\mu(A_0) = \mu(\bigcup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu(A_n) = 0$, contradicting the fact that g is not zero a.e. $\square$

We will not use the following, but it is interesting. It shows that Riemann integration is restricted to a very narrow class of functions, whereas Lebesgue integration is much more general.

PROPOSITION 1.40. *If f is bounded on a compact set $[a,b] \subset \mathbb{R}$, then f is Riemann integrable on $[a,b]$ if and only if f is continuous at a.e. point of $[a,b]$.*

The Lebesgue integral is absolutely continuous in the following sense.

THEOREM 1.41. *If $f \in \mathcal{L}(\Omega)$, then $\int_A |f|\,dx \to 0$ as $\mu(A) \to 0$, where $A \subset \Omega$ is measurable. That is, given $\epsilon > 0$, there is $\delta > 0$ such that*

$$\int_A |f(x)|\,dx \le \epsilon$$

whenever $\mu(A) < \delta$.

Proof. Given $\epsilon > 0$, there is a simple function $s(x)$ such that

$$\int_A |f(x) - s(x)|\, dx \le \epsilon/2 \ ,$$

by the definitionn of the Lebesgue integral. Moreover, by the proof of the existance of $s(x)$, we know that we can take $s(x)$ bounded:

$$|s(x)| \le M(\epsilon)$$

for some $M(\epsilon)$. Then on $A \subset \Omega$ measurable,

$$\int_A |s(x)|\, dx \le \mu(A) M(\epsilon),$$

so if $\mu(A) < \delta \equiv \epsilon/2M(\epsilon)$, then

$$\int_A |f(x)|\, dx \le \int_A |f(x) - s(x)|\, dx + \int_A |s(x)|\, dx \le \epsilon/2 + \epsilon/2 = \epsilon \ .$$

□

The following lemma is easily demonstrated (and left to the reader), but it turns out to be quite useful.

Lemma 1.42 (Chebyshev's Inequality). *If $f \ge 0$ and $\Omega \subset \mathbb{R}^d$ are measurable, then*

$$\mu(\{x \in \Omega : f(x) > \alpha\}) \le \frac{1}{\alpha} \int_\Omega f(x)\, dx$$

for any $\alpha > 0$.

We conclude our overview of Lebesgue measure and integration with the three basic convergence theorems, Fubini's Theorem on integration over product spaces, and the Fundamental Theorem of Calculus, each without proof. For the first three results, assume that $\Omega \subset \mathbb{R}^d$ is measurable.

Theorem 1.43 (Lebesgue's Monotone Convergence Theorem). *If $\{f_n\}_{n=1}^\infty$ is a sequence of measurable functions satisfying $0 \le f_1(x) \le f_2(x) \le \cdots$ for a.e. $x \in \Omega$, then*

$$\lim_{n\to\infty} \int_\Omega f_n(x)\, dx = \int_\Omega \Big(\lim_{n\to\infty} f_n(x) \Big)\, dx \ .$$

Theorem 1.44 (Fatou's Lemma). *If $\{f_n\}_{n=1}^\infty$ is a sequence of nonnegative, measurable functions, then*

$$\int_\Omega \Big(\liminf_{x\to\infty} f_n(x) \Big)\, dx \le \liminf_{n\to\infty} \int_\Omega f_n(x)\, dx \ .$$

Theorem 1.45 (Lebesgue's Dominated Convergence Theorem). *Let $\{f_n\}_{n=1}^\infty$ be a sequence of measurable functions that converge pointwise for a.e. $x \in \Omega$. If there is a function $g \in \mathcal{L}(\Omega)$ such that*

$$|f_n(x)| \le g(x) \quad \textit{for every } n \textit{ and a.e. } x \in \Omega,$$

then

$$\lim_{n\to\infty} \int_\Omega f_n(x)\, dx = \int_\Omega \Big(\lim_{n\to\infty} f_n(x) \Big)\, dx \ .$$

THEOREM 1.46 (Fubini's Theorem). *Let f be measurable on $\mathbb{R}^{n+m}$. If at least one of the integrals*

$$I_1 = \int_{\mathbb{R}^{n+m}} f(x,y)\,dx\,dy \ ,$$
$$I_2 = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x,y)\,dx \right) dy \ ,$$
$$I_3 = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x,y)\,dy \right) dx$$

exists in the Lebesgue sense (i.e., when f is replaced by $|f|$) and is finite, then each exists and $I_1 = I_2 = I_3$.

Note that in Fubini's Theorem, the claim is that the following are equivalent:

(i) $f \in \mathcal{L}(\mathbb{R}^{n+m})$,
(ii) $f(\cdot, y) \in \mathcal{L}(\mathbb{R}^n)$ for a.e. $y \in \mathbb{R}^m$ and $\int_{\mathbb{R}^n} f(x,\cdot)\,dx \in \mathcal{L}(\mathbb{R}^m)$,
(iii) $f(x,\cdot) \in \mathcal{L}(\mathbb{R}^m)$ for a.e. $x \in \mathbb{R}^n$ and $\int_{\mathbb{R}^m} f(\cdot,y)\,dy \in \mathcal{L}(\mathbb{R}^n)$,

and the three full integrals agree. Among other things, f being measurable on $\mathbb{R}^{n+m}$ implies that $f(\cdot, y)$ is measurable for a.e. $y \in \mathbb{R}^m$ and $f(x,\cdot)$ is measurable for a.e. $x \in \mathbb{R}^n$. Note also that we cannot possibly claim anything about *every* $x \in \mathbb{R}^n$ and/or $y \in \mathbb{R}^m$, but only about *almost every* point.

THEOREM 1.47 (Fundamental Theorem of Calculus). *If $f \in \mathcal{L}([a,b])$ and*

$$F(x) = \int_a^x f(t)\,dt \ ,$$

then $F'(x) = f(x)$ for a.e. $x \in [a,b]$. Conversely, if F is differentiable everywhere (not a.e.!) on $[a,b]$ and $F' \in \mathcal{L}([a,b])$, then

$$F(x) - F(a) = \int_a^x F'(t)\,dt$$

for any $x \in [a,b]$.

1.3. The Lebesgue spaces $L_p(\Omega)$

Let $\Omega \subset \mathbb{R}^d$ be measurable and let $0 < p < \infty$. We denote by $L_p(\Omega)$ the class of all measurable functions $f : \Omega \to \mathbb{R}$ (or $\mathbb{C}$) such that

$$\int_\Omega |f(x)|^p\,dx < \infty \ . \tag{1.2}$$

An interesting point arises here. Suppose f and g lie in $L_p(\Omega)$ and that $f(x) = g(x)$ for a.e. $x \in \Omega$. Then as far as integration is concerned, one really cannot distinguish f from g. For example, if $A \subset \Omega$ is measurable, then

$$\int_A |f|^p\,dx = \int_A |g|^p\,dx \ .$$

Thus within the class of $L_p(\Omega)$, f and g are equivalent. This is formalized by modifying the definition of the elements of $L_p(\Omega)$. We declare two measurable functions that are equal a.e. to be equivalent, and define the elements of $L_p(\Omega)$ to be the equivalence classes

$$[f] = \{g : \Omega \to \mathbb{R} \text{ (or } \mathbb{C}) \ : g = f \text{ a.e. on } \Omega\}$$

such that one (and hence all) representative function satisfies (1.2). However, for convenience, we continue to speak of and denote elements of $L_p(\Omega)$ as "functions" which may be modified on a set of measure zero without consequence. For example, $f = 0$ in $L_p(\Omega)$ means only that $f = 0$ a.e. in Ω.

The integral (1.2) arises frequently, so we denote it as

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

and call it the $L_p(\Omega)$-*norm*. (A general definition of *norm* will be given later, and $\|\cdot\|_p$ will be an important example.)

A function $f(x)$ is said to be *bounded on* Ω by $K \in \mathbb{R}$ if $|f(x)| \le K$ for every $x \in \Omega$. We modify this for measurable functions.

Definition. A measurable function $f : \Omega \to \mathbb{C}$ is *essentially bounded* on Ω by K if $|f(x)| \le K$ for a.e. $x \in \Omega$. The infimum of such K is the *essential supremum* of $|f|$ on Ω, and denoted $\operatorname{ess\,sup}_{x\in\Omega} |f(x)|$.

For $p = \infty$, we define $\|f\|_\infty = \operatorname{ess\,sup}_{x\in\Omega} |f(x)|$. Then for all $0 < p \le \infty$,

$$L_p(\Omega) = \{f : \|f\|_p < \infty\} .$$

Proposition 1.48. *If $0 < p \le \infty$, then $L_p(\Omega)$ is a vector space and $\|f\|_p = 0$ if and only if $f = 0$ a.e. in Ω.*

Proof. We first show that $L_p(\Omega)$ is closed under addition. For $p < \infty$, $f, g \in L_p(\Omega)$, and $x \in \Omega$,

$$|f(x) + g(x)|^p \le \Big(|f(x)| + |g(x)|\Big)^p \le 2^p \Big(|f(x)|^p + |g(x)|^p\Big) .$$

Integrating, there obtains $\|f + g\|_p \le 2(\|f\|_p^p + \|g\|_p^p)^{1/p} < \infty$. The case $p = \infty$ is clear.

For scalar multiplication, note that for $\alpha \in \mathbb{R}$ (or $\mathbb{C}$),

$$\|\alpha f\|_p = |\alpha| \; \|f\|_p ,$$

so $f \in L_p(\Omega)$ implies $\alpha f \in L_p(\Omega)$. The remark that $\|f\|_p = 0$ implies $f = 0$ a.e. is clear. □

These spaces are interrelated in a number of ways.

Theorem 1.49 (Hölder's Inequality). *Let $1 \le p \le \infty$ and let q denote the* conjugate exponent *defined by*

$$\frac{1}{p} + \frac{1}{q} = 1 \quad (q = \infty \text{ if } p = 1,\ q = 1 \text{ if } p = \infty).$$

If $f \in L_p(\Omega)$ and $g \in L_q(\Omega)$, then $fg \in L_1(\Omega)$ and

$$\|fg\|_1 \le \|f\|_p \|g\|_q .$$

If $1 < p < \infty$, equality occurs if and only if $|f(x)|^p$ and $|g(x)|^q$ are proportional a.e. in Ω.

Proof. The result is clear if $p = 1$ or $p = \infty$. Suppose $1 < p < \infty$. The function $u : [0, \infty) \to \mathbb{R}$ given by

$$u(t) = \frac{t^p}{p} + \frac{1}{q} - t$$

has minimum value 0, attained only with $t = 1$. For $a, b \geq 0$, let $t = ab^{-q/p}$ to obtain from the previous observation that

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \tag{1.3}$$

with equality if and only if $a^p/b^q = 1$. If $\|f\|_p = 0$ or $\|g\|_q = 0$, then $fg = 0$ a.e. on Ω and the result follows. Otherwise let $a = |f(x)|/\|f\|_p$ and $b = |g(x)|/\|g\|_q$ and integrate over Ω. $\square$

REMARK. The same proof works for sequences $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$. The resulting discrete version of Hölder's inequality is

$$\sum_{n=1}^\infty |a_n b_n| \leq \left(\sum_{n=1}^\infty a_n^p\right)^{1/p} \left(\sum_{n=1}^\infty b_n^q\right)^{1/q} .$$

Inequality (1.3) is extremely useful, and often used when $p = q = 2$, so we make formal note of it below.

PROPOSITION 1.50. *If a and b are nonnegative real numbers, $1 < p < \infty$, and q is the conjugate exponent to p (i.e., $1/p + 1/q = 1$), then*

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} .$$

Moreover, for any $\epsilon > 0$, then there is $C = C(p, \epsilon) > 0$ such that

$$ab \leq \epsilon a^p + C b^q .$$

THEOREM 1.51 (Minkowski's Inequality). *If $1 \leq p \leq \infty$ and f and g are measurable, then*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p .$$

PROOF. If f or $g \notin L_p(\Omega)$, the result is clear, since the right-hand side is infinite. The result is also clear for $p = 1$ or $p = \infty$, so suppose $1 < p < \infty$ and $f, g \in L_p(\Omega)$. Then

$$\begin{aligned} \|f + g\|_p^p = \int_\Omega |f(x) + g(x)|^p \, dx &\leq \int_\Omega |f(x) + g(x)|^{p-1} \Big(|f(x)| + |g(x)|\Big) \, dx \\ &\leq \left(\int_\Omega |f(x) + g(x)|^{(p-1)q} \, dx\right)^{1/q} \Big(\|f\|_p + \|g\|_p\Big) , \end{aligned}$$

by two applications of Hölder's inequality, where $1/p + 1/q = 1$. Since $(p-1)q = p$ and $1/q = (p-1)/p$,

$$\|f + g\|_p^p \leq \|f + g\|_p^{p-1} (\|f\|_p + \|g\|_p) .$$

The integral on the left is finite, so we can cancel terms (unless $\|f + g\|_p = 0$, in which case there is nothing to prove). $\square$

PROPOSITION 1.52. *Suppose $\Omega \subset \mathbb{R}^d$ has finite measure ($\mu(\Omega) < \infty$) and $1 \leq p \leq q \leq \infty$. If $f \in L_q(\Omega)$, then $f \in L_p(\Omega)$ and*

$$\|f\|_p \leq \big(\mu(\Omega)\big)^{1/p - 1/q} \|f\|_q .$$

If $f \in L_\infty(\Omega)$, then

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

If $f \in L_p(\Omega)$ for $1 \le p < \infty$ and there is $K > 0$ such that

$$\|f\|_p \le K \ ,$$

then $f \in L_\infty(\Omega)$ and $\|f\|_\infty \le K$.

We leave the proof of this as an exercise, though the latter two results are nontrivial.

Let $d : L_p(\Omega) \times L_p(\Omega) \to \mathbb{R}$ be defined by

$$d(f,g) = \|f-g\|_p \ .$$

It is easy to verify with Minkowski's inequality that d is a metric. Thus $L_p(\Omega)$ is both a metric space and a vector space. It is an important spoace of functions that arises in many branches of applied mathematics. It has a rich structure involving topological and algebraic concepts and their interplay. It is an example of a more general class of spaces called Banach spaces. We study these in the next chapter.

1.4. Exercises

1. Show that the following define a topology $\mathcal{T}$ on X, where X is any nonempty set.

   (a) $T = \{\emptyset, X\}$. This is called the *trivial* topology on X.

   (b) $T_B = \{\{x\} : x \in X\}$ is a base. This is called the *discrete* topology on X.

   (c) Let $\mathcal{T}$ consist of $\emptyset$ and all subsets of X with finite complements. If X is finite, what topology is this?

2. Let $X = \{a,b\}$ and $\mathcal{T} = \{\emptyset, \{a\}, X\}$. Show directly that there is no metric $d : X \times X \to \mathbb{R}$ that is compatible with the topology. Thus not every topological space is *metrizable.*

3. Prove that if $A \subset X$, then ∂A is closed and

   $$\bar{A} = A^\circ \cup \partial A \ , \qquad A^\circ \cap \partial A = \emptyset \ .$$

   Moreover,

   $$\partial A = \partial A^c = \{x \in X : \text{ every open } E \text{ containing } x \text{ intersects both } A \text{ and } A^c\} \ .$$

4. Prove that if $(X, \mathcal{T})$ is Hausdorff, then every set consisting of a single point is closed. Moreover, limits of sequences are unique.

5. Prove that a set $A \subset X$ is open if and only if, given $x \in A$, there is an open E such that $x \in E \subset A$.

6. Prove that a mapping of X into Y is continuous if and only if the inverse image of every closed set is closed.

7. Prove that if f is continuous and $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} f(x_n) = f(x)$.

8. Suppose that $f(x) = y$. Let $\mathcal{B}_x$ be a base at $x \in X$, and $\mathcal{C}$ a base at $y \in Y$. Prove that f is continuous at x if and only if for each $C \in \mathcal{C}_y$ there is a $B \in \mathcal{B}_x$ such that $B \subset f^{-1}(C)$.

9. Show that every metric space is Hausdorff.

10. Suppose that $F : X \to \mathbb{R}$. Characterize all topologies $\mathcal{T}$ on X that make f continuous. Which is the weakest? Which is the strongest?

11. Construct an infinite open cover of $(0,1]$ that has no finite subcover. Find a sequence in $(0,1]$ that does not have a convergent subsequence.

12. Prove that the continuous image of a compact set is compact.

13. Prove that a one-to-one continuous map of a compact space X onto a Hausdorff space Y is necessarily a homeomorphism.

14. Prove that if $f : X \to \mathbb{R}$ is continuous and X compact, then f takes on its maximum and minimum values.

15. Show that the Borel sets $\mathcal{B}$ is the collection of all sets that can be constructed by a countable number of basic set operations, starting from open sets. The basic set operations consist of taking unions, intersections, or complements.

16. Prove each of the following.

    (a) If $f : \mathbb{R}^d \to \mathbb{R}$ is measurable and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $g \circ f$ is measurable.

    (b) If $\Omega \subset \mathbb{R}^d$ is measurable and $f : \Omega \to \mathbb{R}$ is continuous, than f is measurable.

17. Let $x \in \mathbb{R}^d$ be fixed. Define d_x for any $A \subset \mathbb{R}^d$ by

$$d_x(A) = \begin{cases} 1 & \text{if } x \in A \ , \\ 0 & \text{if } x \notin A \ . \end{cases}$$

    Show that d_x is a measure on the Borel sets $\mathcal{B}$. This measure is called the Dirac or point measure at x.

18. The Divergence Theorem from advanced calculus says that if $\Omega \subset \mathbb{R}^d$ has a smooth boundary and $\mathbf{v} \in (C^1(\bar{\Omega}))^d$ is a vector-valued function, then

$$\int_\Omega \nabla \cdot \mathbf{v}(x)\, dx = \int_{\partial\Omega} \mathbf{v}(x) \cdot \nu(x)\, ds(x) \ ,$$

    where $\nu(x)$ is the outward pointing unit normal vector to Ω for any $x \in \partial\Omega$, and $ds(x)$ is the surface differential (i.e., measure) on $\partial\Omega$. Note that here dx is a d-dimensional measure, and ds is a $(d-1)$-dimensional measure.

    (a) Interpret the formula when $d = 1$ in terms of the Dirac measure.

    (b) Show that for $\phi \in C^1(\bar{\Omega})$,

$$\nabla \cdot (\phi \mathbf{v}) = \nabla\phi \cdot \mathbf{v} + \phi \nabla \cdot \mathbf{v} \ .$$

    (c) Let $\phi \in C^1(\bar{\Omega})$ and apply the Divergence Theorem to the vector $\phi\mathbf{v}$ in place of $\mathbf{v}$. We call this new formula *integration by parts*. Show that it reduces to ordinary integration by parts when $d = 1$.

19. Prove that if $f \in \mathcal{L}(\Omega)$ and $g : \Omega \to \mathbb{R}$, where g and Ω are measurable and g is bounded, then $fg \in \mathcal{L}(\Omega)$.

20. Construct an example of a sequence of nonnegative measurable functions from $\mathbb{R}$ to $\mathbb{R}$ that shows that strict inequality can result in Fatou's Lemma.

21. Let

$$f_n(x) = \begin{cases} \dfrac{1}{n} \ , & |x| \le n \ , \\ 0 \ , & |x| > n \ . \end{cases}$$

Show that $f_n(x) \to 0$ uniformly on $\mathbb{R}$, but

$$\int_{-\infty}^{\infty} f_n(x)\,dx = 2\ .$$

Comment on the applicability of the Dominated Convergence Theorem.

22. Let

$$f(x,y) = \begin{cases} 1\ , & 0 \le x - y \le 1\ , \\ -1\ , & 0 \le y - x \le 1\ , \\ 0\ , & \text{otherwise.} \end{cases}$$

Show that

$$\int_0^{\infty} \left(\int_0^{\infty} f(x,y)\,dx \right) dy \ne \int_0^{\infty} \left(\int_0^{\infty} f(x,y)\,dy \right) dx\ .$$

Comment on the applicability of Fubini's Theorem.

23. Suppose that f is integrable on $[a,b]$, and define

$$F(x) = \int_a^x f(t)\,dt\ .$$

Prove that F is continuous on $[a,b]$. (In fact, $F' = f$ a.e., but it is more involved to prove this.)

24. Suppose that $\Omega \subset \mathbb{R}^d$ has finite measure and $1 \le p \le q \le \infty$.

(a) Prove that if $f \in L_q(\Omega)$, then $f \in L_p(\Omega)$ and

$$\|f\|_p \le (\mu(\Omega))^{1/p-1/q} \|f\|_q\ .$$

(b) Prove that if $f \in L_\infty(\Omega)$, then

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

(c) Prove that if $f \in L_p(\Omega)$ for all $1 \le p < \infty$, and there is $K > 0$ such that $\|f\|_p \le K$, then $f \in L_\infty(\Omega)$ and $\|f\|_\infty \le K$.

CHAPTER 2

Normed Linear Spaces and Banach Spaces

2.1. Introduction

Functional Analysis grew out of the late 19th century study of differential and integral equations arising in physics, but it emerged as a subject in its own right in the first part of the 20th century. Thus functional analysis is a genuinely 20th century subject, often the first one a student meets in analysis. For the first sixty or seventy years of this century, functional analysis was a major topic within mathematics, attracting a large following among both pure and applied mathematicians. Lately, the pure end of the subject has become the purview of a more restricted coterie who are concerned with very difficult and often quite subtle issues. On the other hand, the applications of the basic theory and even of some of its finer elucidations has grown steadily, to the point where one can no longer intelligently read papers in much of numerical analysis, partial differential equations and parts of stochastic analysis without a working knowledge of functional analysis. Indeed, the basic structures of the theory arises in many other parts of mathematics and its applications.

Our aim in the first section of this course is to expound the elements of the subject with an eye especially for aspects that lend themselves to applications.

We begin with a formal development as this is the most efficient path.

Definition. Let X be a vector space over $\mathbb{R}$ or $\mathbb{C}$. We say X is a *normed linear space* (NLS for short) if there is a mapping

$$\|\cdot\| : X \to \mathbb{R}^+ = [0,\infty)$$

called the *norm* on X, satisfying the following set of rules which apply to $x, y \in X$ and $\lambda \in \mathbb{R}$ or $\mathbb{C}$:

$$\begin{aligned}
&\|\lambda x\| = |\lambda|\,\|x\| \ , \\
&\|x\| = 0 \ \text{ if and only if } \ x = 0 \ , \\
&\|x + y\| \le \|x\| + \|y\| \qquad \text{(triangle inequality)}.
\end{aligned}$$

In situations where more than one NLS is under consideration, it is often convenient to write $\|\cdot\|_X$ for the norm on the space X to indicate which norm is connoted.

Examples. Consider $\mathbb{R}^d$ or $\mathbb{C}^d$, with the usual Euclidean length of a vector x denoted $|x|$. If we define, for $x \in \mathbb{R}^d$ or $\mathbb{C}^d$,

$$\|x\| = |x| \ ,$$

then $(\mathbb{R}^d, \|\cdot\|)$ or $(\mathbb{C}^d, \|\cdot\|)$ is a NLS.

Let p lie in the range $[1,\infty)$ and define the real vector spaces and norms

$$\ell_p = \left\{ x = \{x_n\}_{n=1}^\infty : x_n \in \mathbb{R} \ \text{ and } \ \|x\|_{\ell_p} = \left(\sum_{n=1}^\infty |x_n|^p \right)^{1/p} < +\infty \right\} .$$

These spaces are NLS's over $\mathbb{R}$. If the sequences are allowed to have complex values, then ℓ_p is a complex NLS. If $p = \infty$, define

$$\ell_\infty = \left\{ x = \{x_n\}_{n=1}^\infty : \|x\|_{\ell_\infty} = \sup_n |x_n| < +\infty \right\} .$$

Let $c_0 \subset \ell_\infty$ be the linear subspace defined as

$$c_0 = \left\{ \{x_n\}_{n=1}^\infty : \lim_{n\to+\infty} x_n = 0 \right\} .$$

Another interesting subspace is

$$f = \left\{ \begin{array}{lcl} \{x_n\}_{n=1}^\infty & : & x_n = 0 \text{ except for a finite} \\ & & \text{number of values of } n \end{array} \right\} .$$

These normed linear spaces are related to each other; indeed if $1 \le p < \infty$,

$$f \subseteq \ell_p \subseteq c_0 \subseteq \ell_\infty .$$

Let a and b be real numbers, $a < b$, with $a = -\infty$ or $b = +\infty$ allowed as possible values. Then

$$C([a,b]) = \left\{ f : [a,b] \to \mathbb{R} \text{ or } \mathbb{C} : f \text{ is continuous and } \sup_{x\in[a,b]} |f(x)| < +\infty \right\} .$$

For $f \in C([a,b])$, let

$$\|f\| = \sup_{x\in[a,b]} |f(x)| .$$

Here the vector space structure is given by pointwise multiplication and addition; that is

$$(f+g)(x) = f(x) + g(x)$$

and

$$(\lambda f)(x) = \lambda f(x) .$$

Remark. In a good deal of the theory developed here, it will not matter for the outcome whether the NLS's are real or complex vector spaces. When this point is moot, we will often write $\mathbb{F}$ rather than $\mathbb{R}$ or $\mathbb{C}$. The reader should understand when the symbol $\mathbb{F}$ appears that it stands for either $\mathbb{R}$ or for $\mathbb{C}$, and the discussion at that juncture holds for both.

A NLS X is *finite dimensional* if it is finite dimensional as a vector space, which is to say there is a finite collection $\{x_j\}_{j=1}^N \subset X$ such that any $x \in X$ can be written as a linear combination of the $\{x_j\}_{j=1}^N$, *viz.*

$$x = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_N x_N ,$$

where the λ_i are scalars (member of the ground field $\mathbb{F}$). Otherwise, X is called *infinite dimensional*. Interest here is mainly in infinite-dimensional spaces.

If X is a NLS and $\|\cdot\|$ and $\|\cdot\|_1$ are two norms on X, they are said to be *equivalent norms* if there exist constants $c, d > 0$ such that

$$c\|x\| \le \|x\|_1 \le d\|x\| \tag{2.1}$$

for all $x \in X$. It is a fundamental fact that on a finite-dimensional NLS, any pair of norms is equivalent, whereas this is not the case in infinite dimensional spaces.

For example, let p lie in the range $1 \le p < \infty$. If $x = \{x_i\}_{i=1}^\infty \in \ell_p$, let $\|\cdot\|$ and $\|\cdot\|_1$ be given by

$$\|x\| = \|x\|_{\ell_p} = \left(\sum_{i=1}^\infty |x_i|^p \right)^{1/p}$$

and

$$\|x\|_1 = \|x\|_{\ell_\infty} = \sup_{i\geq 1} |x_i| \ .$$

These both define norms on ℓ_p, but they are not equivalent.

Let $(X, \|\cdot\|)$ be a NLS. Then X is a *metric space* if we define a metric d on X by

$$d(x,y) = \|x-y\| \ .$$

To see this, just note the following: for $x, y, z \in X$,

$$\begin{aligned}
d(x,x) &= \|x-x\| = \|0\| = 0 \ , \\
0 = d(x,y) &= \|x-y\| \implies x-y=0 \\
&\qquad\qquad\quad \implies x=y \ , \\
d(x,y) &= \|x-y\| = \|-(y-x)\| \\
&= |-1|\,\|y-x\| = d(y,x) \ , \\
d(x,y) &= \|x-y\| = \|x-z+z-y\| \\
&\leq \|x-z\| + \|z-y\| = d(x,z) + d(z,y) \ .
\end{aligned}$$

Consequently, the concepts of elementary topology are available in any NLS. In particular, we may talk about open sets and closed sets in a NLS.

A set $U \subset X$ is *open* if for each $x \in U$, there is an $r > 0$ (depending on x in general) such that

$$B_r(x) = \{y \in X : d(y,x) < r\} \subset U \ .$$

The set $B_r(x)$ is referred to as the (open) ball of radius r about x. A set $F \subset X$ is closed if $X \smallsetminus F = \{y \in X,\ y \notin F\}$ is open. As with any metric space, F is closed if it is sequentially closed. That is, a set F is closed if, whenever $\{x_n\}_1^\infty \subseteq F$ and $x_n \to x$ for the metric, then it must be the case that $x \in F$. If $\|\cdot\|$ and $\|\cdot\|_1$ are two equivalent norms on a NLS X, then the collections O and O_1 of open sets induced by these two norms as just outlined are the same. Thus topologically, $(X, \|\cdot\|)$ and $(X, \|\ \|_1)$ are indistinguishable.

Recall that a sequence $\{x_n\}_{n=1}^\infty$ in a metric space (X, d) is called a *Cauchy sequence* if

$$\lim_{n,m\to\infty} d(x_n, x_m) = 0 \ ;$$

or equivalently, given $\varepsilon > 0$, there is an $N = N(\varepsilon)$ such that if $n, m \geq N$, then

$$d(x_n, x_m) \leq \varepsilon \ .$$

A metric space is called *complete* if every Cauchy sequence converges. A NLS $(X, \|\cdot\|)$ that is complete as a metric space is called a *Banach space* after the Polish mathematician Stefan Banach who was a pioneer in the subject.

Examples. $\mathbb{R}^d$ and $\mathbb{C}^d$ are complete as we learned in advanced calculus or elementary analysis. The spaces ℓ_p, $1 \leq p \leq \infty$ are complete, though this requires proof. However, if we take the space ℓ_1 and equip it with the ℓ_∞-norm, this is a NLS, but *not* a Banach space.

To check this, first note that ℓ_1 is a linear subspace of ℓ_∞. Indeed, if $x = (x_1, x_2, \dots) \in \ell_1$, then

$$|x|_{\ell_\infty} = \sup_{i\geq 1} |x_i| \leq \sum_{j=1}^\infty |x_j| = |x|_{\ell_1} \ .$$

Hence ℓ_1 with the ℓ_∞-norm is a NLS. To see it is not complete, consider the following sequence. Define $\{y_k\}_{k=1}^\infty \subset \ell_1$ by

$$y_k = (y_{k,1}, y_{k,2}, \dots) = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{k}, 0, 0, \dots\right) ,$$

$k = 1, 2, 3, \dots$. Then $\{y_k\}_{k=1}^\infty$ is Cauchy in the ℓ_∞-norm. For if $k \geq m$, then

$$|y_k - y_m|_{\ell_\infty} \leq \frac{1}{m+1} .$$

If ℓ_1 were complete in the ℓ_∞-norm, then $\{y_k\}_{k=1}^\infty$ would converge to some element $z \in \ell_1$. Thus we would have that

$$|y_k - z|_{\ell_\infty} \to 0$$

as $k \to +\infty$. But, for $j \geq 1$,

$$|y_{k,j} - z_j| \leq |y_k - z|_{\ell_\infty}$$

where $y_{k,j}$ and z_j are the j^{th}-components of y_k and z, respectively. In consequence, it is seen that $z_j = 1/j$ for all $j \geq 1$. However, the element

$$z = \left(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{k}, \frac{1}{k+1}, \dots\right)$$

does not lie in ℓ_1, a contradiction.

If X is a linear space over $\mathbb{F}$ and d is a metric on X induced from a norm on X, then for all $x, y, a \in X$ and $\lambda \in \mathbb{F}$ that

$$d(x + a, y + a) = d(x, y) \qquad \text{and} \qquad d(\lambda x, \lambda y) = |\lambda| d(x, y) . \tag{2.2}$$

Question. Suppose X is a linear space over $\mathbb{R}$ or $\mathbb{C}$ and d is a metric on X satisfying (2.2). Is it necessarily the case that there is a norm $\|\cdot\|$ on X such that $d(x, y) = \|x - y\|$?

Definition. A set C in a linear space X over $\mathbb{F}$ is *convex* if whenever $x, y \in C$, then so also is

$$tx + (1 - t)y$$

whenever $0 \leq t \leq 1$.

Proposition 2.1. *Suppose $(X, \|\cdot\|)$ is a NLS and $r > 0$. For any $x \in X$, $B_r(x)$ is convex.*

Proof. Let $y, z \in B_r(x)$ and $t \in [0, 1]$ and compute as follows:

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

Thus, $B_r(x)$ is convex. $\square$

Corollary 2.2. *If $p < 1$, then $\|\cdot\|_p$ is not a norm on ℓ_p.*

To prove this, show the unit ball $B_1(0)$ is not convex. It is also easy to see the triangle inequality does not always hold.

One reason vector spaces are so important and ubiquitous is that they are the natural domain of definition for linear maps, and the latter pervade mathematics and its applications. Remember, a linear map is one that commutes with addition and scalar multiplication, so that

$$T(x+y) = T(x) + T(y) ,$$
$$T(\lambda x) = \lambda T(x) ,$$

for $x, y \in X$, $\lambda \in \mathbb{F}$.

On the other hand, the natural mappings between topological spaces, and metric spaces in particular, are the continuous maps. If (X, d) and (Y, ρ) are two metric spaces and $f : X \to Y$ is a function, then f is *continuous* if for any $x \in X$ and $\varepsilon > 0$, there exists a $\delta = \delta(x, \varepsilon) > 0$ such that

$$d(x, y) \le \delta \text{ implies } \rho(f(x), f(y)) \le \varepsilon .$$

If $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ are NLS's, then they are simultaneously linear spaces and metric spaces. Thus one might expect the collection

$$B(X, Y) = \{T : X \to Y : T \text{ is linear and continuous}\} \tag{2.3}$$

to be an interesting class of mappings that are consistent with both the algebraic and metric structures of the underlying spaces. Continuous linear mappings between NLS's are often called *bounded operators* or *bounded linear operators* or *continuous linear operators*.

Proposition 2.3. *Let X and Y be NLS's and $T : X \to Y$ a linear map. The following are equivalent:*

(a) *T is continuous,*
(b) *T is continuous at some point,*
(c) *T is bounded on bounded sets.*

Proof. (a $\implies$ b) Trivial.

(b $\implies$ c) Suppose T is continuous at $x_0 \in X$. Let M be a bounded set in X and let $R > 0$ be such that $M \subset B_R(0)$. By continuity at x_0, there is a $\delta = \delta(1, x_0) > 0$ such that

$$\|x - x_0\|_X \le \delta \text{ implies } \|Tx - Tx_0\|_Y \le 1 . \tag{2.4}$$

But by linearity $Tx - Tx_0 = T(x - x_0)$. Thus, (2.4) is equivalent to the condition

$$\|y\|_X \le \delta \text{ implies } \|Ty\|_Y \le 1 .$$

Hence, it follows readily that if $\|y\|_X \le R$, then

$$\|Ty\|_Y = \Big\|\frac{R}{\delta} T\Big(\frac{\delta}{R} y\Big)\Big\|_Y = \frac{R}{\delta}\Big\|T\Big(\frac{\delta}{R} y\Big)\Big\|_Y \le \frac{R}{\delta}$$

since

$$\Big\|\frac{\delta}{R} y\Big\| \le \delta .$$

It then transpires that

$$\sup\{\|Tx\|_Y : x \in M\} \le \frac{R}{\delta} < +\infty .$$

(c $\implies$ a) It is supposed that T is linear and bounded on bounded sets. In particular, there is an $R > 0$ such that

$$T(B_1(0)) \subset B_R(0) .$$

Let $\varepsilon > 0$ be given and let $\delta = \varepsilon / R$. Suppose $\|x - x_0\|_X \le \delta$. Then by homogeneity,

$$\Big\|\frac{1}{\delta}(x - x_0)\Big\|_X \le 1 ,$$

whence

$$\begin{array}{ccc}\left\|T\Big(\frac{1}{\delta}(x-x_0)\Big)\right\|_Y & \leq & R \\ \| & & \\ \frac{1}{\delta}\|Tx-Tx_0\|_Y\ . & & \end{array}$$

Thus, if $\|x-x_0\|\leq\delta$, then

$$\|Tx-Tx_0\|_Y \leq R\delta = \varepsilon\ .$$

Therefore T is continuous at x_0, and x_0 was an arbitrary point in X. □

Let X, Y be NLS's and let $T \in B(X,Y)$ be a continuous linear operator from X to Y. We know that T is therefore bounded on any bounded set of X, so the quantity

$$\|T\| = \|T\|_{B(X,Y)} = \sup_{x\in B_1(0)} \|Tx\|_Y \tag{2.5}$$

is finite. The notation makes it clear that this mapping $\|\cdot\|_{B(X,Y)} : B(X,Y) \to [0,\infty)$ is expected to be a norm. There are several things to check.

First, $B(X,Y)$ is a vector space in its own right if we define $S+T$ and λT by

$$(S+T)(x) = Sx + Tx\ ,$$

and

$$(\lambda S)(x) = \lambda Sx\ .$$

Proposition 2.4. *Let X and Y be NLS's. The formula* (2.5) *defines a norm on $B(X,Y)$. Moreover, if $T\in B(X,Y)$, then*

$$\|T\| = \sup_{\|x\|=1}\|Tx\|_Y = \sup_{x\neq 0}\frac{\|Tx\|_Y}{\|x\|_X}\ . \tag{2.6}$$

If Y is a Banach space, then so is $B(X,Y)$ with this norm.

Proof. If T is the zero map, then clearly $\|T\|=0$. On the other hand, if $\|T\|=0$, then T vanishes on the closed unit ball. For any $x\in X$, $x\neq 0$, write $x = \|x\|\frac{1}{\|x\|}x = \|x\|y$. Then y is in the closed unit ball, so $T(y)=0$. Then $T(x) = \|x\|T(y) = 0$; thus $T\equiv 0$. Plainly, by definition of scalar multiplication

$$\|\lambda T\| = \sup_{B_1(0)}\|(\lambda T)(x)\|_Y = |\lambda|\sup_{B_1(0)}\|Tx\|_Y = |\lambda|\,\|T\|\ .$$

The triangle inequality is just as simple:

$$\begin{aligned}\|T+S\| &= \sup_{x\in B_1(0)}\|(T+S)(x)\|_Y = \sup_{x\in B_1(0)}\|Tx+Sx\|_Y \\ &\leq \sup_{x\in B_1(0)}\Big\{\|Tx\|_Y + \|Sx\|_Y\Big\} \\ &\leq \sup_{x\in B_1(0)}\|Tx\|_Y + \sup_{x\in B_1(0)}\|Sx\|_Y = \|T\|+\|S\|\ .\end{aligned}$$

Thus $(B(X,Y), \|\cdot\|_{B(X,Y)})$ is indeed a NLS.

The alternative formulas for the norm expressed in (2.6) are straightforward to deduce. Notice that the last formula makes it obvious that for all $x\in X$ and $T\in B(X,Y)$,

$$\|Tx\|_Y \leq \|T\|_{B(X,Y)}\|x\|_X\ , \tag{2.7}$$

an inequality that will find frequent use.

The more interesting fact is that $B(X,Y)$ is complete if we only assume Y is complete. This simple result has far-reaching consequences. To establish this point, suppose $\{T_n\}_{n=1}^\infty$ is a Cauchy sequence in $B(X,Y)$. We must show it converges in $B(X,Y)$. Let $x \in X$ and consider the sequence $\{T_n x\}_{n=1}^\infty$ in Y. Because of (2.7), it follows that

$$\|T_n x - T_m x\|_Y \le \|T_n - T_m\|_{B(X,Y)} \|x\|_Y \ ,$$

and thus $\{T_n x\}_{n=1}^\infty$ is seen to be Cauchy in Y. As Y is a Banach space, $\{T_n x\}_{n=1}^\infty$ must converge to some element of Y that depends upon x of course; call this element Tx. There is thus established a correspondence

$$x \longmapsto Tx$$

between X and Y. We claim it is a continuous linear correspondence, whence $T \in B(X,Y)$. It is further asserted that $T_n \to T$ in $B(X,Y)$.

First note that

$$\begin{aligned} T(x+y) &= \lim_{n\to\infty} T_n(x+y) = \lim_{n\to\infty} \{T_n x + T_n y\} \\ &= \lim_{n\to\infty} T_n x + \lim_{n\to\infty} T_n y = Tx + Ty \ . \end{aligned}$$

Similarly, $T(\lambda x) = \lambda Tx$ for $x \in X$ and $\lambda \in \mathbb{F}$. Thus T is a linear map. Also, T is a bounded map. First, remark that $\{T_n\}_{n=1}^\infty$, being Cauchy, must be a bounded sequence. For there is an N such that if $n \ge N$, then

$$\|T_n - T_N\| \le 1 \ ,$$

say. By the triangle inequality, this means

$$\|T_n\| \le \|T_N\| + 1 \ ,$$

for $n \ge N$. The initial segment, $\{T_1, T_2, \dots, T_{N-1}\}$ of the sequence is bounded since it is finite, say $\|T_j\| \le K$ for $1 \le j \le N-1$. It therefore transpires that

$$\|T_k\| \le \max\{K, \|T_N\| + 1\} = M \ ,$$

say, for all k. From this it follows at once that T is a bounded operator; for if $x \in X$, then

$$\|Tx\|_Y = \lim_{n\to\infty} \|T_n x\|_Y \le \limsup_{n\to\infty} \|T_n\|_{B(X,Y)} \|x\|_X \le M\|x\| \ .$$

Finally, we check that $T_n \to T$ in $B(X,Y)$. Let $x \in B_1(0)$ in X and observe that

$$\begin{aligned} \|Tx - T_n x\|_Y &= \lim_{m\to\infty} \|T_m x - T_n x\|_Y = \lim_{m\to\infty} \|(T_m - T_n)x\| \\ &\le \limsup_{m\to\infty} \|T_m - T_n\|_{B(X,Y)} \|x\|_X \le \varepsilon(n) \ . \end{aligned}$$

Since x was an arbitrary element in $B_1(0)$, this means

$$\|T - T_n\|_{B(X,Y)} \le \varepsilon(n)$$

and because $\{T_k\}_{k=1}^\infty$ is Cauchy, $\varepsilon(n) \to 0$ as $n \to +\infty$. □

DEFINITION. Let X be a NLS over $\mathbb{F}$. The *dual space* X^* of X is the Banach space $B(X,\mathbb{F})$. The elements of X^* are called *bounded linear functionals* on X.

REMARK. The dual space is complete because $\mathbb{R}$ and $\mathbb{C}$ are complete. At first glance, it is not so clear that X^* is interesting to study; it might even reduce to the trivial vector space if X is large. It will turn out to be quite a fruitful object to understand, however.

Attention is now turned to the three principal results in the elementary theory of Banach spaces. These theorems will find frequent use in many parts of the course.

2.2. Hahn-Banach Theorems

The Hahn-Banach theorems enable us to extend linear functionals defined on a subspace to the entire space. The theory begins with the case when the underlying field $\mathbb{F} = \mathbb{R}$ is real, and the first crucial lemma enables us to extend by a single dimension. The main theorem then follows from this result and an involved induction argument. The corresponding result over $\mathbb{C}$ follows as a corollary from an important observation relating complex and real linear functionals. In the case of a NLS, we can even extend the functional continuously. But first a definition.

Definition. Let X be a vector space over $\mathbb{F}$. We say that $p : X \to [0,\infty)$ is *sublinear* if it satisfies for any $x, y \in X$ and $\lambda \geq 0$

$$\begin{aligned} p(\lambda x) &= \lambda\, p(x) && \text{(positive homogeneous)},\\ p(x+y) &\leq p(x) + p(y) && \text{(triangle inequality)}. \end{aligned}$$

If p also satisfies for any $x \in X$ and $\lambda \in \mathbb{F}$

$$p(\lambda x) = |\lambda|\, p(x) \ ,$$

then p is said to be a *seminorm*.

Thus a sublinear function p is a seminorm if and only if it satisfies the stronger homogeneity property $p(\lambda x) = |\lambda| p(x)$ for any $\lambda \in \mathbb{F}$, and a seminorm p is a norm if and only if $p(x) = 0$ implies that $x = 0$.

Lemma 2.5. *Let X be a vector space over $\mathbb{R}$ and let $Y \subset X$ be a linear subspace such that $Y \neq X$. Let p be sublinear on X and $f : Y \to \mathbb{R}$ be a linear map such that*

$$f(y) \leq p(y) \tag{2.8}$$

for all $y \in Y$. For a given $x_0 \in X \setminus Y$, let

$$\tilde{Y} = \operatorname{span}\{Y, x_0\} = Y + \mathbb{R}\, x_0 = \{y + \lambda x_0 : y \in Y,\ \lambda \in \mathbb{R}\} \ .$$

Then there exists a linear map $\tilde{f} : \tilde{Y} \to \mathbb{R}$ such that

$$\tilde{f}|_Y = f \quad \textit{and} \quad -p(-x) \leq \tilde{f}(x) \leq p(x) \tag{2.9}$$

for all $x \in \tilde{Y}$.

Proof. We need only find $\tilde{f}$ such that $\tilde{f}(x) \leq p(x)$, since then we also have $-\tilde{f}(x) = \tilde{f}(-x) \leq p(-x)$.

Suppose there was such an $\tilde{f}$. What would it have to look like? Let $\tilde{y} = y + \lambda x_0 \in \tilde{Y}$. Then, by linearity,

$$\tilde{f}(\tilde{y}) = \tilde{f}(y) + \lambda \tilde{f}(x_0) = f(y) + \lambda\alpha \ , \tag{2.10}$$

where $\alpha = \tilde{f}(x_0)$ is some real number. Therefore, such an $\tilde{f}$, were it to exist, is completely determined by α. Conversely, a choice of α determines a well-defined linear mapping. Indeed, if

$$\tilde{y} = y + \lambda x_0 = y' + \lambda' x_0 \ ,$$

then

$$y - y' = (\lambda' - \lambda) x_0 \ .$$

The left-hand side lies in Y, while the right-hand side can lie in Y only if $\lambda' - \lambda = 0$. Thus $\lambda = \lambda'$ and then $y = y'$. Hence the representation of x in the form $y + \lambda x_0$ is unique and so a choice of $\tilde{f}(x_0) = \alpha$ determines a unique linear mapping by using the formula (2.10) as its definition.

It remains to be seen whether it is possible to choose α so that (2.9) holds. This amounts to asking that for all $y \in Y$ and $\lambda \in \mathbb{R}$,

$$f(y) + \lambda\alpha = \tilde{f}(y + \lambda x_0) \leq p(y + \lambda x_0) \ . \tag{2.11}$$

Now, (2.11) is true for $\lambda = 0$ by the hypothesis (2.8). If $\lambda \neq 0$, write $y = -\lambda x$, or $x = -\frac{1}{\lambda}y$. Then, (2.11) becomes

$$-\lambda(f(x) - \alpha) \leq p(-\lambda(x - x_0))$$

or, when $\lambda < 0$,

$$f(x) - \alpha \leq p(x - x_0) \ ,$$

and, when $\lambda > 0$,

$$-(f(x) - \alpha) \leq p(-(x - x_0)) \ ,$$

for all $x \in Y$. This is the same as the two-sided inequality

$$-p(x_0 - x) \leq f(x) - \alpha \leq p(x - x_0)$$

or,

$$f(x) - p(x - x_0) \leq \alpha \leq f(x) + p(x_0 - x) \ . \tag{2.12}$$

Thus any choice of α that respects (2.12) for all $x \in Y$ leads via (2.10) to a linear map $\tilde{f}$ with the desired property. Is there such an α? Let

$$a = \sup_{x \in Y} f(x) - p(x - x_0)$$

and

$$b = \inf_{x \in Y} f(x) + p(x_0 - x) \ .$$

If it is demonstrated that $a \leq b$, then there certainly is such an α and any choice in the non-empty interval $[a, b]$ will do. But, a calculation shows that for $x, y \in Y$,

$$f(x) - f(y) = f(x - y) \leq p(x - y) \leq p(x - x_0) + p(x_0 - y) \ ,$$

on account of (2.8) and the triangle inequality. In consequence, we have

$$f(x) - p(x - x_0) \leq f(y) + p(x_0 - y) \ ,$$

and this holds for any $x, y \in Y$. Fixing y, we see that

$$\sup_{x \in Y} f(x) - p(x - x_0) \leq f(y) + p(x_0 - y) \ .$$

As this is valid for every $y \in Y$, it must be the case that

$$a = \sup_{x \in Y} f(x) - p(x - x_0) \leq \inf_{y \in Y} f(y) + p(x_0 - y) = b \ .$$

The result is thereby established. $\square$

We now want to successively extend f to all of X, one dimension at a time. We can do this trivially if $X \setminus Y$ is finite dimensional. If $X \setminus Y$ were to have a countable vector space basis, we could use ordinary induction. However, not many interesting NLS's have a countable vector space basis. We therefore need to consider the most general case of a possibly uncountable vector space basis, and this requires that we use what is known as *transfinite induction.*

We begin with some terminology.

Definition. For a set S, an *ordering*, denoted by $\leq$, is a binary relation such that:

(a) $x \leq x$ for every $x \in S$ (reflexivity);
(b) If $x \leq y$ and $y \leq x$, then $x = y$ (antisymmetry);
(c) If $x \leq y$ and $y \leq z$, then $x \leq z$ (transitivity).

A set S is *partially ordered* if S has an ordering that may apply only to certain pairs of elements of S, that is, there may be x and y in S such that neither $x \leq y$ nor $y \leq x$ holds. In that case, x and y are said to be *incomparable;* otherwise they are *comparable.* A *totally ordered set* or *chain* C is a partially ordered set such that every pair of elements in C are comparable.

Lemma 2.6 (Zorn's Lemma). *Suppose S is a nonempty, partially ordered set. Suppose that every chain $C \subset S$ has an* upper bound; *that is, there is some $u \in S$ such that*

$$x \leq u \quad \text{for all } x \in C \ .$$

Then S has at least one maximal element; *that is, there is some $m \in S$ such that for any $x \in S$,*

$$m \leq x \quad \implies m = x \ .$$

This lemma follows from the *Axiom of Choice*, which states that given any set S and any collection of its subsets, we can choose a single element from each subset. In fact, Zorn's lemma implies the Axiom of Choice, and is therefore equivalent to it. Since the proof takes us deeply into logic and far afield from Functional Analysis, we accept Zorn's lemma as an Axiom of set theory and proceed.

Theorem 2.7 (Hahn-Banach Theorem for Real Vector Spaces). *Suppose that X is a vector space over $\mathbb{R}$, Y is a linear subspace, and p is sublinear on X. If f is a linear functional on Y such that*

$$f(x) \leq p(x) \tag{2.13}$$

for all $x \in Y$, then there is a linear functional F on X such that

$$F|_Y = f$$

(i.e., F is a linear extension *of f) and*

$$-p(-x) \leq F(x) \leq p(x)$$

for all $x \in X$.

Proof. Let S be the set of all linear extensions g of f, defined on a vector space $D(g)$, and satisfying the property $g(x) \leq p(x)$ for all $x \in D(g)$. Since $f \in S$, S is not empty. We define a partial ordering on S by $g \leq h$ means that h is an extension of g. More precisely, $g \leq h$ means that $D(g) \subset D(h)$ and $g(x) = h(x)$ for all $x \in D(g)$.

For any chain $C \in S$, let

$$D = \bigcup_{g \in C} D(g) \ ,$$

which is easily seen to be a vector space since C is a chain. Define for $x \in D$

$$g_C(x) = g(x)$$

for any $g \in C$ such that $x \in D(g)$. Again, since C is a chain, g_C is well defined. Moreover, it is linear and $D(g_C) = D$. Hence, g_C is in S and it is an upper bound for the chain C.

We can therefore apply Zorn's Lemma to conclude that S has at least one maximal element F. By definition, F is a linear extension satisfying $F(x) \le p(x)$ for all $x \in D(F)$. It remains to show that $D(F) = X$. If not, there is some nonzero $x \in X \setminus D(F)$, and by the previous extension result, we can extend F to $\tilde{F}$ on $D(F) + \mathbb{R}x$. This contradicts the maximality of F, so F is a linear extension satisfying our desired properties. $\square$

Theorem 2.8 (Hahn-Banach Theorem for General Vector Spaces). *Suppose that X is a vector space over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$), Y is a linear subspace, and p is a seminorm on X. If f is a linear functional on Y such that*

$$|f(x)| \le p(x) \tag{2.14}$$

for all $x \in Y$, then there is a linear functional F on X such that

$$F|_Y = f$$

(i.e., F is a linear extension *of f) and*

$$|F(x)| \le p(x)$$

for all $x \in X$.

Proof. Write f in terms of its real and imaginary parts, *viz.* $f = g + ih$, where g and h are real-valued. Clearly $g(y + z) = g(y) + g(z)$ and $h(y + z) = h(y) + h(z)$. If $\lambda \in \mathbb{R}$, then

$$\begin{aligned} f(\lambda x) &= g(\lambda x) + ih(\lambda x) \\ \| \\ \lambda f(x) &= \lambda g(x) + i\lambda h(x)\ . \end{aligned}$$

Taking real and imaginary parts in this relation and combining with the fact that g and h commute with addition shows them both to be real linear. Moreover, g and h are intimately related. To see this, remark that for $x \in Y$,

$$\begin{aligned} f(ix) &= if(x) = ig(x) - h(x) = -h(x) + ig(x) \\ \| \\ g(ix) &+ ih(ix)\ . \end{aligned}$$

Taking the real part of this relation leads to

$$g(ix) = -h(x)\ ,$$

so that, in fact,

$$f(x) = g(x) - ig(ix)\ . \tag{2.15}$$

Since g is the real part of f, clearly for $x \in Y$,

$$|g(x)| \le |f(x)| \le p(x) \tag{2.16}$$

by assumption. Thus g is a real-linear map defined on Y, considered as a vector subspace of X over $\mathbb{R}$. Because of (2.16), g satisfies the hypotheses of Theorem 2.7, so we obtain an extension G of g such that G is an $\mathbb{R}$-linear map of X into $\mathbb{R}$ which is such that

$$|G(x)| \le p(x)$$

for all $x \in X$. Use (2.15) to define F:

$$F(x) = G(x) - iG(ix) \ .$$

It is to be shown that F is a $\mathbb{C}$-linear extension of f to X and, moreover, for all $x \in X$,

$$|F(x)| \le p(x) \ . \tag{2.17}$$

First we check that F is $\mathbb{C}$-linear. As it is $\mathbb{R}$-linear, it suffices to show $F(ix) = iF(x)$. But this is true since

$$F(ix) = G(ix) - iG(-x) = G(ix) + iG(x) = i\big(G(x) - iG(ix)\big) = iF(x) \ .$$

Inequality (2.17) holds for the following reason. Let $x \in X$ and write $F(x) = re^{i\theta}$ for some $r \ge 0$. Then, we have

$$r = |F(x)| = e^{-i\theta}F(x) = F(e^{-i\theta}x) = G(e^{-i\theta}x) \le p(e^{-i\theta}x) = p(x) \ ,$$

since $F(e^{-i\theta}x)$ is real. $\square$

Corollary 2.9 (Hahn-Banach Theorem for Normed Linear Spaces). *Let X be a NLS over $\mathbb{F}$ ($\mathbb{R}$ or $\mathbb{C}$) and let Y be a linear subspace. Let $f \in Y^*$ be a continuous linear functional on Y. Then there is an $F \in X^*$ such that*

$$F|_Y = f$$

and

$$\|F\|_{X^*} = \|f\|_{Y^*} \ .$$

Proof. Simply apply the Hahn-Banach Theorem to f, using seminorm

$$p(x) = \|f\|_{Y^*}\|x\|_X \ .$$

We leave the details to the reader. $\square$

2.3. Applications of Hahn-Banach

Corollary 2.10. *Let X be a NLS and $x_0 \ne 0$ in X. Then there is an $f \in X^*$ such that*

$$\|f\|_{X^*} = 1 \quad \textit{and} \quad f(x_0) = \|x_0\| \ .$$

Proof. Let $Z = \mathbb{F}x_0 = \operatorname{span}\{x_0\}$. Define h on Z by

$$h(\lambda x_0) = \lambda\|x_0\| \ .$$

Then $h : Z \to \mathbb{F}$ and h has norm one on Z since for $x \in Z$, say $x = \lambda x_0$,

$$|h(x)| = |h(\lambda x_0)| = \big|\lambda\|x_0\|\,\big| = \|\lambda x_0\| = \|x\| \ .$$

By the Hahn-Banach Theorem, there exists $f \in X^*$ such that $f|_Z = h$ and $\|f\| = \|h\| = 1$. $\square$

Corollary 2.11. *Let X be a NLS and $x \in X$. There exists an $f \in X^*$ such that*

$$f(x) = \|f\|_{X^*}\|x\| \ .$$

The proof is similar to that above.

Corollary 2.12. *Let X be a NLS and $x_0 \in X$. Then*

$$\|x_0\| = \sup_{\substack{f\in X^* \\ f\neq 0}} \frac{|f(x_0)|}{\|f\|_{X^*}} = \sup_{\substack{f\in X^* \\ \|f\|=1}} |f(x_0)| \ .$$

Proof. In any event, we always have

$$\frac{|f(x_0)|}{\|f\|_{X^*}} \le \frac{\|f\|_{X^*}\|x_0\|_X}{\|f\|_{X^*}} = \|x_0\| \ ,$$

and consequently

$$\sup_{\substack{f\in X^* \\ f\neq 0}} \frac{|f(x_0)|}{\|f\|_{X^*}} \le \|x_0\| \ .$$

On the other hand, by Corollary 2.11, there is an $\tilde f \in X^*$ such that $\tilde f(x_0) = \|\tilde f\| \, \|x_0\|$. It follows that

$$\sup_{\substack{f\in X^* \\ f\neq 0}} \frac{|f(x_0)|}{\|f\|_{X^*}} \ge \frac{|\tilde f(x_0)|}{\|\tilde f\|_{X^*}} = \|x_0\| \ . \qquad \square$$

Proposition 2.13. *Let X be a NLS. Then X^* separates points in X.*

Proof. Let $x_1, x_2 \in X$, with $x_1 \neq x_2$. Then $x_2 - x_1 \neq 0$, so by Corollary 1.8, there is an $f \in X^*$ so that

$$f(x_2 - x_1) \neq 0 \ .$$

Since f is linear, this means

$$f(x_2) \neq f(x_1) \ ,$$

which is the desired conclusion. $\square$

Corollary 2.14. *Let X be a NLS and $x_0 \in X$ such that $f(x_0) = 0$ for all $f \in X^*$. Then $x_0 = 0$.*

Proof. This follows from either of the last two results. $\square$

Lemma 2.15 (Mazur Separation Lemma 1). *Let X be a NLS, Y a linear subspace of X and $w \in X \smallsetminus Y$. Suppose*

$$d = \operatorname{dist}(w, Y) = \inf_{z\in Y} \|w - z\|_X > 0 \ .$$

Then there exists $f \in X^*$ such that $\|f\|_{X^*} \le 1$,

$$f(w) = d \quad \text{and} \quad f(z) = 0 \quad \text{for all} \quad z \in Y \ .$$

Proof. As before, any element $x \in Z = Y + \mathbb{F}w$ has a unique representation in the form $x = y + \lambda w$. Define $g : Z \to \mathbb{F}$ by

$$g(y + \lambda w) = \lambda d \ .$$

It is easy to see g is $\mathbb{F}$-linear and that $\|g\|_{Z^*} \le 1$. The latter is true since, if $x \in Z$, $x = y + \lambda w \neq 0$, then if $\lambda = 0$, $x \in Y$ and so $|f(x)| = 0 \le 1$, whereas if $\lambda \neq 0$, then

$$f\left(\frac{y+\lambda w}{\|y+\lambda w\|}\right) = \frac{\lambda}{\|y+\lambda w\|}\ d = \frac{1}{\|\frac{1}{\lambda}y + w\|}\ d \ .$$

Since $\frac{1}{\lambda}y = -z \in Y$, it follows that

$$\left\| \frac{1}{\lambda} y + w \right\| \geq d \ .$$

In consequence, we have

$$f\left(\frac{y + \lambda w}{\|y + \lambda w\|} \right) \leq \frac{d}{d} = 1 \ .$$

Use the Hahn-Banach Theorem to extend f to an $F \in X^*$ without increasing its norm. The functional F meets the requirements in view. $\square$

Proposition 2.16. *Let X be a Banach space and X^* its dual. If X^* is separable, then so is X.*

Proof. Let $\{f_n\}_{n=1}^\infty$ be a countable dense subset of X^*. Let $\{x_n\}_1^\infty \subset X$, be such that

$$\|x_n\| = 1 \ \text{ and } \ |f_n(x_n)| \geq \tfrac{1}{2}\|f_n\| \ , \qquad n = 1, 2, \dots \ .$$

Such elements $\{x_n\}_{n=1}^\infty$ exist by definition of the norm on X^*.

Let $\mathcal{D}$ be the countable set

$$\mathcal{D} = \left\{ \begin{array}{c} \text{all finite linear combinations of the } \{x_j\}_{j=1}^\infty \\ \text{with rational coefficients} \end{array} \right\}$$

We claim that $\mathcal{D}$ is dense in X. If $\mathcal{D}$ is *not* dense in X, then there is an element $w_0 \in X \smallsetminus \overline{\mathcal{D}}$. The point w_0 is at positive distance from $\overline{\mathcal{D}}$, for if not, there is a sequence $\{z_n\}_{n=1}^\infty \subset \overline{\mathcal{D}}$ such that $z_n \to w_0$. As $\overline{\mathcal{D}}$ is closed, this means $z \in \overline{\mathcal{D}}$ and that contradicts the choice of w_0.

From Lemma 2.15, there is an $f \in X^*$ such that

$$f|_{\overline{\mathcal{D}}} = 0 \ \text{ and } \ f(w_0) = d = \inf_{z \in \overline{\mathcal{D}}} \|z - w_0\|_X \ .$$

Since $f \in X^*$, there is a subsequence $\{f_{n_k}\}_{k=1}^\infty \subset X^*$ such that $f_{n_k} \xrightarrow{X^*} f$, by density. In consequence,

$$\begin{aligned} \|f - f_{n_k}\|_{X^*} &\geq |(f - f_{n_k})(x_{n_k})| \\ &= |f_{n_k}(x_{n_k})| \geq \tfrac{1}{2}\|f_{n_k}\|_{X^*} \ . \end{aligned}$$

Hence $\|f_{n_k}\|_{X^*} \to 0$ as $k \to \infty$, and this means $f = 0$, a contradiction since $f(w) = d > 0$. $\square$

We can also use the Hahn-Banach Theorem to distinguish sets that are not strictly subspaces, as long as the linear geometry is respected. The next two lemmas consider convex sets.

Lemma 2.17 (Mazur Separation Lemma 2). *Let X be a NLS, C a closed, convex subset of X such that $\lambda x \in C$ whenever $x \in C$ and $|\lambda| \leq 1$ (we say that such a set C is* balanced*). For any $w \in X \smallsetminus C$, there exists $f \in X^*$ such that $|f(x)| \leq 1$ for all $x \in C$ and $f(w) > 1$.*

Proof. Let $B \in X$ be an open ball containing w that does not intersect C. Define the *Minkowski functional* $p : X \to [0, \infty)$ by

$$p(x) = \inf\{t > 0 : x/t \in C + B\} \ .$$

Since $0 \in C$, $p(x)$ is indeed finite for every $x \in X$ (i.e., eventually every point can be contracted at least into the ball $0 + B$). Moreover, $p(x) \leq 1$ for $x \in C$, but $p(w) > 1$.

We claim that p is a seminorm. First, given $x \in X$, $\lambda \in \mathbb{F}$, and $t > 0$, the condition $\lambda x/t \in C + B$ is equivalent to $|\lambda| x/t \in (|\lambda|/\lambda)(C + B) = C + B$, since C and B are balanced. Thus

$$p(\lambda x) = p(|\lambda| x) = |\lambda| p(x) \ .$$

Second, if $x, y \in X$ and we choose any $s > 0$ and $r > 0$ such that $x/r \in C + B$ and $y/s \in C + B$, then the convex combination

$$\frac{r}{s+r}\frac{x}{r} + \frac{s}{s+r}\frac{y}{s} = \frac{x+y}{s+r} \in C + B \ ,$$

and so we conclude that

$$p(x+y) \leq p(x) + p(y) \ .$$

Now let $Y = \mathbb{F}w$ and define on Y the linear functional

$$f(\lambda w) = \lambda p(w),$$

so $f(w) = p(w) > 1$. Now

$$|f(\lambda w)| = |\lambda|\, p(w) = p(\lambda w) \ ,$$

so the Hahn-Banach Theorem gives us a linear extension with the property that

$$|f(x)| \leq p(x)$$

that is, $|f(x)| \leq 1$ for $x \in C \subset C + B$, as required. Finally, f is bounded on B, so it is continuous. $\square$

Not all convex sets are balanced, so we have the following lemma. We can no longer require that the entire linear functional be well behaved when $\mathbb{F} = \mathbb{C}$, but only its real part.

LEMMA 2.18 (Separating Hyperplane Theorem). *Let A and B be disjoint, nonempty, convex sets in a NLS X.*

(a) *If A is open, then there is $f \in X^*$ and $\gamma \in \mathbb{R}$ such that*

$$\mathrm{Re} f(x) \leq \gamma \leq \mathrm{Re} f(y) \quad \forall x \in A \ , \ y \in B \ .$$

(b) *If both A and B are open, then there is $f \in X^*$ and $\gamma \in \mathbb{R}$ such that*

$$\mathrm{Re} f(x) < \gamma < \mathrm{Re} f(y) \quad \forall x \in A \ , \ y \in B \ .$$

(c) *If A is compact and B is closed, then there is $f \in X^*$ and $\gamma \in \mathbb{R}$ such that*

$$\mathrm{Re} f(x) < \gamma < \mathrm{Re} f(y) \quad \forall x \in A \ , \ y \in B \ .$$

PROOF. It is sufficient to prove the result for field $\mathbb{F} = \mathbb{R}$. Then if $\mathbb{F} = \mathbb{C}$, we have a continuous, real-linear functional g satisfying the separation result. We construct $f \in X^*$ by using (2.15):

$$f(x) = g(x) - ig(ix) \ .$$

So we consider now only the case of a real field $\mathbb{F} = \mathbb{R}$.

For (a), fix $-w \in A - B = \{x - y : x \in A, y \in B\}$ and let

$$C = A - B + \{w\} \ ,$$

which is an open, convex neighborhood of 0 in X. Moreover, $w \notin C$, since A and B are disjoint. Define the subspace $Y = \mathbb{R}w$ and the linear functional $g : Y \to \mathbb{R}$ by

$$g(tw) = t \ .$$

Now let $p : X \to [0, \infty)$ be the Minkowski functional for C,

$$p(x) = \inf\{t > 0 : x/t \in C\} \ .$$

We saw in the previous proof that p is sublinear (it is not necessarily a seminorm, since C may not be balanced, but it does satisfy the triangle inequality and positive homogeneity). Since $w \notin C$, $p(w) \geq 1$ and $g(y) \leq p(y)$ for $y \in Y$, so we use the Hahn-Banach Theorem for real functionals (Theorem 2.7) to extend g to X linearly. Now $g \le 1$ on C, so also $g \ge -1$ on $-C$, and we conclude that $|g| \le 1$ on $C \cap (-C)$, which is a neighborhood of 0. Thus g is bounded, and so continuous.

If $x \in A$ and $y \in B$, then $a - b + w \in C$, so

$$1 \ge g(a - b + w) = g(a) - g(b) + g(w) = g(a) - g(b) + 1 \ ,$$

which implies that $g(a) \le g(b)$, and the result follows with $\gamma = \sup_{a \in A} g(A)$.

For (b), we use the previous construction. It is left to the reader to show that $g(A)$ is an open subset of $\mathbb{R}$, since g is linear and A is open. Now both $g(A)$ and $g(B)$ are open subsets that can intersect only in one point, so they must be disjoint.

For (c), consider $S \equiv B - A$. Since A is compact, we claim that S is closed. So suppose there are points $x_n \in S$ such that $x_n = b_n - a_n$ with $b_n \in B$ and $a_n \in A$ and $x_n \to x$ in X. But since A is compact, there is a subsequence (still denoted by a_n for convenience), such that $a_n \to a \in A$. But then $b_n = x_n + a_n \to x + a \equiv b \in B$, since B is closed. But this implies that $x \in S$, and the claim follows.

Since $0 \notin S$, there is some open convex set $U \in X$ containing 0 such that $U \cap S$ is empty. Let $A' = A + \frac{1}{2}U$ and $B' = B - \frac{1}{2}U$. Then A' and B' are disjoint, convex, open sets, and so (b) gives a functional with the desired properties, which hold also for the subsets $A \subset A'$ and $B \subset B'$. $\square$

2.4. The Embedding of X into its Double Dual

Let X be a NLS and X^* its dual space. Since X^* is a Banach space, it has a dual space X^{**} which is sometimes referred to as the *double dual* of X. There is a natural construction whereby X may be viewed as a subspace of X^{**} that is described now.

For any $x \in X$, define $T_x = [x] \in X^{**}$ as follows: if $f \in X^*$, then

$$T_x(f) = [x](f) = f(x) \ . \tag{2.18}$$

First, lets check that this is an element of X^{**}. We need to see that $[x]$ is a bounded linear map on X^*. Let $f, g \in X^*$, $\lambda \in \mathbb{F}$ and compute as follows:

$$\begin{aligned} [x](f+g) &= (f+g)(x) \\ &= f(x) + g(x) \\ &= [x](f) + [x](g) \ , \end{aligned}$$

and

$$\begin{aligned} [x](\lambda f) &= (\lambda f)(x) \\ &= \lambda f(x) \\ &= \lambda [x](f) \ . \end{aligned}$$

Thus $[x]$ is a linear map of X^* into $\mathbb{F}$. It is bounded since, by Corollary 2.12,

$$\|[x]\|_{X^{**}} = \sup_{\substack{f \in X^* \\ f \ne 0}} \frac{|[x](f)|}{\|f\|_{X^*}} = \|x\| \ .$$

Thus, not only is $[x]$ bounded, but its norm in X^{**} is the same as the norm of x in X. Thus we may view X as a linear subspace of X^{**}, and in this guise, X is faithfully represented in X^{**}.

DEFINITION. Let (M, d) and (N, ρ) be two metric spaces and $f : M \to N$. The function f is called an *isometry* if f preserves distances, which is to say

$$\rho(f(x), f(y)) = d(x, y) \ .$$

The spaces M and N are called *isometric* if there is a surjective isometry $f : M \to N$.

Metric spaces that are isometric are indistinguishable as metric spaces. If the metric spaces are NLS's $(X, \| \ \|_X)$ and $(Y, \| \ \|_Y)$ and $T : X \to Y$ is a linear isometry, then $T(X)$ may be identified with X.

In this terminology, the correspondence $F : X \to X^{**}$ given by

$$F(x) = [x]$$

is an isometry. A NLS space X is called *reflexive* if F is surjective. In this case, X is necessarily a Banach space.

2.5. The Open Mapping Theorem

The second of the three major principles of elementary functional analysis is the Open Mapping Theorem (or equivalently the Closed Graph Theorem). The third is the principle of uniform boundedness (Banach-Steinhaus theorem). Both of these rely on the following theorem of Baire.

THEOREM 2.19 (Baire Category Theorem). *Let X be a complete metric space. Then the intersection of any countable collection of dense open sets in X is dense in X.*

PROOF. Let $\{V_j\}_{i=1}^{\infty}$ be a countable collection of dense open sets. Let W be any non-empty open set in X. It is required to show that if $V = \bigcap_{j=1}^{\infty} V_j$, then $V \cap W \neq \emptyset$. Since V_1 is dense, $W \cap V_1$ is a non-empty open set. Thus there is an $r_1 > 0$ and an $x_1 \in W$, and without loss of generality, $r_1 < 1$, such that

$$\overline{B_{r_1}(x_1)} \subset W \cap V_1 \ .$$

Similarly, V_2 is open and dense, hence there is an x_2 and an r_2 with $0 < r_2 < 1/2$ such that

$$\overline{B_{r_2}(x_2)} \subset V_2 \cap B_{r_1}(x_1) \ .$$

Inductively, we determine x_n, r_n with $0 < r_n < 1/n$ such that

$$\overline{B_{r_n}(x_n)} \subset V_n \cap B_{r_{n-1}}(x_{n-1}) \ , \qquad n = 2, 3, 4, \dots \ .$$

Consider the sequence $\{x_n\}_{n=1}^{\infty}$ just generated. If $i, j \geq n$, then by construction

$$x_i, x_j \in B_{r_n}(x_n) \implies d(x_i, x_j) \leq \frac{2}{n} \ .$$

This shows that $\{x_i\}_{n=1}^{\infty}$ is a Cauchy sequence. As X is complete, there is an x for which $x_i \to x$ as $i \to +\infty$. Because $x_i \in \overline{B_{r_n}(x_n)}$ for $i > n$, it follows that $x \in \overline{B_{r_n}(x_n)}$, $n = 1, 2, \dots$. Hence $x \in V_n$, $n = 1, 2, \dots$. Clearly, since $x \in \overline{B_{r_1}(x_1)} \subset W$, $x \in W$ also. Hence

$$x \in W \cap \bigcap_{n=1}^{\infty} V_n \ ,$$

and the proof is complete. □

COROLLARY 2.20. *The intersection of countably many dense open subsets of a complete metric space is non-empty.*

Definition. A set A is called *nowhere dense* if $\mathrm{Int}(\bar{A}) = \emptyset$. A set is called *first category* if it is a countable union of nowhere dense sets. Otherwise, it is called *second category*.

Corollary 2.21. *A complete metric space is second category.*

Proof. If $X = \bigcup_{j=1}^{\infty} M_j$ where each M_j is nowhere dense, then $X = \bigcup_{j=1}^{\infty} \bar{M}_j$, so by deMorgan's law,

$$\emptyset = \bigcap_{j=1}^{\infty} (X \setminus \bar{M}_j) \ .$$

But, for each j, $X \setminus \bar{M}_j$ is open and dense since, by Prop. 1.7,

$$\overline{X \setminus \bar{M}_j} = X \setminus \mathrm{Int}(\bar{M}_j) = X \ .$$

This contradicts Baire's theorem. ◻

Theorem 2.22 (Open-Mapping Principle). *Let X and Y be Banach spaces and let $T : X \to Y$ be a bounded linear surjection. Then T is an open mapping, i.e., T maps open sets to open sets.*

Proof. It is required to demonstrate that if U is open in X, then $T(U)$ is open in Y. If $y \in T(U)$, we must show $T(U)$ contains an open set about y. Suppose it is known that there is an $r > 0$ for which $T(B_1(0)) \supset B_r(0)$. Let $x \in U$ be such that $Tx = y$ and let $t > 0$ be such that $B_t(x) \subset U$. Then, we see that

$$\begin{aligned} T(U) &\supset T(B_t(x)) = T(tB_1(0) + x) \\ &= tT(B_1(0)) + Tx \supset tB_r(0) + y \\ &= B_{rt}(y) \ . \end{aligned}$$

As $rt > 0$, y is an interior point of $T(U)$ and the result would be established. Thus attention is concentrated on showing that $T(U) \supset B_r(0)$ for some $r > 0$ when $U = B_1(0)$.

We continue to write U for $B_1(0)$. Since T is onto,

$$Y = \bigcup_{k=1}^{\infty} T(kU) \ .$$

Since Y is a complete metric space, at least one of the sets $T(kU)$, $k = 1, 2, \dots$, is not nowhere dense. Hence there is a non-empty open set W_1 such that

$$W_1 \subseteq \overline{T(kU)} \quad \text{for some } k \geq 1 \ .$$

Multiplying this inclusion by $1/2k$ yields a non-empty open set $W = \frac{1}{2k} W_1$ included in $\overline{T(\frac{1}{2}U)}$. Hence there is a $y_0 \in Y$ and an $r > 0$ such that

$$B_r(y_0) \subset W \subset \overline{T(\tfrac{1}{2}U)} \ .$$

But then, it must be the case that

$$B_r(0) = B_r(y_0) - y_0 \subset B_r(y_0) - B_r(y_0) \subset \overline{T(\tfrac{1}{2}U)} - \overline{T(\tfrac{1}{2}U)} \subset \overline{T(U)} \ . \tag{2.19}$$

The latter inclusion is very nearly the desired conclusion. It is only required to remove the overbar on the right-hand side. Note that since multiplication by a non-zero constant is a homeomorphism, (2.19) implies that for any $s > 0$,

$$B_{rs}(0) \subset \overline{T(sU)} \ . \tag{2.20}$$

Fix $y \in B_r(0)$ and an ε in $(0,1)$. Since $T(U) \cap B_r(0)$ is dense in $B_r(0)$, there exists $x_1 \in U$ such that

$$\|y - Tx_1\|_Y < \tfrac{1}{2}\gamma \ ,$$

where $\gamma = r\varepsilon$. We proceed by mathematical induction. Let $n \geq 1$ and suppose $x_1, x_2, \ldots, x_n$ have been chosen so that

$$\|y - Tx_1 - Tx_2 - \cdots - Tx_n\|_Y < \left(\tfrac{1}{2}\right)^n \gamma \ . \tag{2.21}$$

Let $z = y - (Tx_1 + \cdots + Tx_n)$. Then $\|z\| < (1/2)^n \gamma$, so because of (2.20), there is an x_{n+1} with

$$\|x_{n+1}\| < \left(\tfrac{1}{2}\right)^n \frac{\gamma}{r} = \left(\tfrac{1}{2}\right)^n \varepsilon \tag{2.22}$$

and

$$\|z - Tx_{n+1}\| < \left(\tfrac{1}{2}\right)^{n+1} \gamma \ .$$

So the induction proceeds and (2.21) and (2.22) hold for all $n \geq 1$.

Now because of (2.22), we know that $\sum_{j=1}^n x_j = S_n$ is Cauchy. Hence, there is an $x \in X$ so that $S_n \to x$ as $n \to +\infty$. Clearly

$$\|x\| \leq \sum_{j=1}^{\infty} \|x_j\| < 1 + \sum_{2}^{\infty} \left(\tfrac{1}{2}\right)^{n+1} \varepsilon = 1 + \varepsilon \ .$$

By continuity of T, $TS_n \to Tx$ as $n \to +\infty$. By (2.21), $TS_n \to y$ as $n \to \infty$. Hence $Tx = y$. Thus we have shown that

$$T((1+\varepsilon)U) \supset B_r(0) \ ,$$

or, what is the same,

$$T(U) \supset B_{r/1+\varepsilon}(0) \ .$$

That establishes the result. $\square$

Corollary 2.23. *Let X, Y be Banach spaces and a T bounded, linear surjection that is also an injection. Then T^{-1} is continuous.*

Proof. This follows since $(T^{-1})^{-1} = T$ is open, hence T^{-1} is continuous. $\square$

A closely related result is the Closed-Graph Theorem. If X, Y are sets and $f : X \to Y$ a function, say defined on a subset $D \subset X$, the *graph of* f is the set

$$\operatorname{graph}(f) = \{(x, y) \in X \times Y : x \in D \text{ and } y = f(x)\} \ .$$

It is a subset of the Cartesian product $X \times Y$.

Proposition 2.24. *Let X be a topological space, Y a Hausdorff space, and $f : X \to Y$ continuous. Then* graph(f) *is closed in $X \times Y$.*

Proof. Let $U = X \times Y \setminus \operatorname{graph}(f)$. Claim U is open. Fix $(x_0, y_0) \in U$, so that $y_0 \neq f(x_0)$. Because Y is Hausdorff, there exist open sets V and W with $y_0 \in V$, $f(x_0) \in W$ and $V \cap W = \emptyset$. Since f is continuous, $f^{-1}(W)$ is open in X. Thus, the open set $f^{-1}(W) \times V$ lies in U. $\square$

Question. Is the last result true if we omit the hypothesis that Y is Hausdorff?

In general, if $f : X \to Y$ and graph(f) is closed, it is not implied that f is continuous. However, in special circumstances, the reverse conclusion is implied.

Definition. Let X and Y be NLS's and let D be a linear subspace of X. Suppose $T : D \to Y$ is linear. Then T is a *closed operator* if graph(T) is a closed subset of $X \times Y$.

Since both X and Y are metric spaces, graph(T) being closed means exactly that if $\{x_n\}_{n=1}^{\infty} \subset D$ with

$$x_n \xrightarrow{X} x \text{ and } Tx_n \xrightarrow{Y} y ,$$

then it follows that $x \in D$ and $y = Tx$.

THEOREM 2.25 (Closed Graph Theorem). *Let X and Y be Banach spaces and $T : X \to Y$ linear. Then T is continuous iff T is closed.*

PROOF. T continuous implies graph(T) is closed on account of Proposition 2.24, since a Banach space is Hausdorff.

Suppose graph(T) to be closed. Then graph(T) is a closed linear subspace of the Banach space $X \times Y$. Hence graph(T) is a Banach space in its own right with the graph norm

$$\|(x, Tx)\| = \|x\|_X + \|Tx\|_Y .$$

Consider the continuous projections Π_1 and Π_2 on $X \times Y$ given by

$$\Pi_1(x, y) = x \text{ and } \Pi_2(x, y) = y .$$

If these are restricted to the subspace graph(T), the following situation obtains:

$$\begin{tikzcd}
& X \times Y \arrow[d, dash] & \\
& \mathrm{graph}(T) \arrow[dl, "\Pi_1"'] \arrow[dr, "\Pi_2"] & \\
X & & Y
\end{tikzcd}$$

The mapping Π_1 is a one-to-one, continuous linear map of the Banach space graph(T) onto X. By the Open Mapping Theorem,

$$\Pi_1^{-1} : X \to \mathrm{graph}(T)$$

is continuous. But then

$$T = \Pi_2 \circ \Pi_1^{-1} : X \to Y$$

is continuous since it is the composition of continuous maps. □

COROLLARY 2.26. *Let X and Y be Banach spaces and D a linear subspace of X. Let $T : D \to Y$ be a closed linear operator. Then T is bounded if and only if D is a closed subspace of X.*

PROOF. As before, if T is closed and linear, then graph(T) is a closed linear subspace of the Banach space $X \times Y$. Hence graph(T) is a Banach space.

If D is closed, it is a Banach space, so the closed graph theorem applied to $T : D \to Y$ shows T to be continuous.

Conversely, suppose T is bounded as a map from D to Y. Let $\{x_n\}_{n=1}^{\infty} \subset D$ and suppose $x_n \to x$ in X. Since T is bounded, it follows that $\{Tx_n\}_{n=1}^{\infty}$ is a Cauchy sequence; for

$$\|Tx_n - Tx_m\| \le \|T\| \, \|x_n - x_m\| \to 0$$

as $n, m \to \infty$. Since Y is complete, there is a $y \in Y$ such that $Tx_n \to y$. But since T is closed, we infer that $x \in D$ and $y = Tx$. In particular, D has all its limit points, so D is closed. □

EXAMPLE. Closed does not imply bounded in general, even for linear operators. Take $X = C(0,1)$ with the max norm. Let $Tf = f'$ for $f \in D = C^1(0,1)$. Consider T as a mapping of D into X.

T is not bounded. Let $f_n(x) = x^n$. Then $\|f_n\| = 1$ for all n, but $Tf_n = nx^{n-1}$ so $\|Tf_n\| = n$.

T is closed. Let $\{f_n\}_{n=1}^\infty \subset D$ and suppose $f_n \xrightarrow{X} f$ and $f_n' \to g$. Then, by the Fundamental Theorem of Calculus,

$$f_n(t) = f_n(0) + \int_0^t f_n'(\tau)\,d\tau$$

for $n = 1, 2, \dots$. Taking the limit of this equation as $n \to \infty$ yields

$$f(t) = f(0) + \int_0^t g(\tau)\,d\tau \ ,$$

so $g = f'$, by another application of the Fundamental Theorem of Calculus.

2.6. Uniform Boundedness Principle

The third basic result in Banach space theory is the Banach-Steinhauss Theorem, also known as the Principle of Uniform Boundedness.

THEOREM 2.27 (Uniform Boundedness Principle). *Let X be a Banach space, Y a NLS and $\{T_\alpha\}_{\alpha \in I} \subset B(X,Y)$ a collection of bounded linear operators from X to Y. Then one of the following two conclusions must obtain: either*

(a) *there is a constant M such that for all $\alpha \in I$,*

$$\|T_\alpha\|_{B(X,Y)} \le M \ ,$$

or

(b) *there is an $x \in X$ such that*

$$\sup_{\alpha \in I} \|T_\alpha x\| = +\infty \ .$$

PROOF. Define the function $\varphi : X \to [0,\infty]$ by

$$\varphi(x) = \sup_{\alpha \in I} \|T_\alpha x\| \ ,$$

for $x \in X$. For $n = 1, 2, 3, \dots$, let

$$V_n = \{x \in X : \varphi(x) > n\} \ .$$

For each $\alpha \in I$, the map φ_α defined by

$$\varphi_\alpha(x) = \|T_\alpha x\|$$

is continuous on X since it is the composition of two continuous maps. Thus the sets

$$\{x : \|T_\alpha x\| > n\} = \varphi_\alpha^{-1}((n,\infty))$$

are open, and consequently,

$$V_n = \bigcup_{\alpha \in I} \varphi_\alpha^{-1}((n,\infty))$$

is a union of open sets, so is itself open. Each V_n is either dense in X or it is not. If for some N, V_N is *not* dense in X, then there is an $r > 0$ and an $x_0 \in X$ such that

$$B_r(x_0) \cap V_N = \emptyset \ .$$

Therefore, if $x \in B_r(x_0)$, then $\varphi(x) \le N$; thus, if $\|z\| < r$, then for all $\alpha \in I$,

$$\|T_\alpha(x_0 + z)\| \le N \ .$$

Hence if $\|z\| \le r$, then for all $\alpha \in I$,

$$\begin{aligned}\|T_\alpha(z)\| &\le \|T_\alpha(z + x_0)\| + \|T_\alpha(x_0)\| \\ &\le N + \|T_\alpha x_0\| \le 2N \ .\end{aligned}$$

In consequence, we have

$$\sup_{\alpha \in I} \|T_\alpha\| \le \frac{2N}{r}$$

and so condition (a) holds.

On the other hand, if all the V_n are dense, then they are all dense and open. By Baire's Theorem,

$$\bigcap_{n=1}^{\infty} V_n$$

is non-empty. Let $x \in \bigcap_{n=1}^{\infty} V_n$. Then, for all $n = 1, 2, 3, \ldots$, $\varphi(x) > n$, and so it follows that $\varphi(x) = +\infty$. $\square$

2.7. Weak Convergence

There are weaker notions of sequential convergence than that induced by the norm on a NLS. Some natural ones play an interesting and helpful role in numerical analysis and the theory of partial differential equations.

Definition. Let X be a NLS and $\{x_n\}_{n=1}^{\infty}$ a sequence in X. We say that $\{x_n\}_{n=1}^{\infty}$ *converges weakly* to $x \in X$ if

$$f(x_n) \to f(x)$$

for all $f \in X^*$. We write $x_n \rightharpoonup x$ or $x_n \xrightarrow{w}$ for weak convergence. Let $\{f_n\}_{n=1}^{\infty}$ be a sequence in X^* and $f \in X^*$. We say that f_n *converges weak-$*$* if for each $x \in X$

$$f_n(x) \to f(x) \ .$$

We write $f_n \xrightarrow{w^*} f$ to indicate weak-$*$ convergence.

Proposition 2.28. *Let X be a NLS and $\{x_n\}_{n=1}^{\infty}$ a sequence from X. If $\{x_n\}_{n=1}^{\infty}$ converges weakly, then its weak limit is unique and $\{\|x_n\|_X\}_{n=1}^{\infty}$ is bounded. If $\{f_n\}_{n=1}^{\infty} \subset X^*$ converges weak-$*$, then its weak-$*$ limit is unique. If in addition X is a Banach space, then $\{\|f_n\|_{X^*}\}_{n=1}^{\infty}$ is bounded.*

Proof. Suppose $x_n \rightharpoonup x$ and $x_n \rightharpoonup y$. That means that for any $f \in X^*$,

$$\begin{array}{ccc} f(x_n) & \longrightarrow & f(x) \\ \downarrow & & \\ f(y) & & \end{array}$$

as $n \to \infty$. Consequently $f(x) = f(y)$ for all $f \in X^*$, which means $x = y$ by the Hahn-Banach Theorem.

Fix an $f \in X^*$. Then the sequence $\{f(x_n)\}_{n=1}^{\infty}$ is bounded in $\mathbb{F}$, say

$$|f(x_n)| \le C_f \quad \text{for all} \quad n$$

since $\{f(x_n)\}_{n=1}^\infty$ converges. View x_n as the evaluation map $E_{x_n} \in X^{**}$. In this context, the last condition amounts to

$$|E_{x_n}(f)| \le C_f$$

for all n. Thus we have a collection of bounded linear maps $\{E_{x_n}\}_{n=1}^\infty$ in $X^{**} = B(X^*, \mathbb{F})$ which are bounded at each point of their domain X^*. By the Uniform Boundedness Principle, which can be applied since X^* is a Banach space, we must have

$$\sup_n \|E_{x_n}\|_{X^{**}} \le C \ .$$

But by the Hahn-Banach Theorem,

$$\|E_{x_n}\|_{X^{**}} = \|x_n\|_X \ .$$

The conclusions for weak-$*$ convergence are left to the reader. $\square$

PROPOSITION 2.29. *Let X be a NLS and $\{x_n\}_{n=1}^\infty \subset X$. If $x_n \rightharpoonup x$, then $\|x\| \le \liminf_{n\to\infty} \|x_n\|$.*

PROOF. Exercise. $\square$

We have actually defined new topologies on X and X^* by these notions of weak convergence.

DEFINITION. Suppose X is a NLS with dual X^*. The *weak topology* on X is the smallest topology on X such that each $f \in X^*$ is continuous. The *weak-$*$ topology* on X^* is the smallest topology on X^* making continuous each evaluation map $E_x : X^* \to \mathbb{F}$, $x \in X$ (defined by $E_x(f) = f(x)$).

A basic open set containing zero in the weak topology of X is of the form

$$U = \{x \in X : |f_i(x)| < \varepsilon_i,\ i = 1, \dots, n\} = \bigcap_{i=1}^n f_i^{-1}(B_{\varepsilon_i}(0))$$

for some n, $\varepsilon_i > 0$, and $f_i \in X^*$. Similarly for the weak-$*$ topology of X^*, a basic open set containing zero is of the form

$$V = \{f \in X^* : |f(x_i)| < \varepsilon_i,\ i = 1, \dots, n\} = \bigcap_{i=1}^n E_{x_i}^{-1}(B_{\varepsilon_i}(0))$$

for some n, $\varepsilon_i > 0$, and $x_i \in X$. The rest of the topology is given by translations and unions of these. If X is infinite dimensional, these topologies are not compatible with any metric, so some care is warrented. That our limit processes arise from these topologies is given by the following.

PROPOSITION 2.30. *Suppose X is a NLS with dual X^*. Let $x \in X$ and $\{x_n\}_{n=1}^\infty \subset X$. Then x_n converges to x in the weak topology if and only if $x_n \overset{w}{\rightharpoonup} x$ (i.e., $f(x_n) \to f(x)$ in $\mathbb{F}$ for every $f \in X^*$). Moreover, if $f \in X^*$ and $\{f_n\}_{n=1}^\infty \subset X^*$, then f_n converges to f in the weak-$*$ topology if and only if $f_n \xrightarrow{w^*} f$ (i.e., $f_n(x) \to f(x)$ in $\mathbb{F}$ for every $x \in X$).*

PROOF. If x_n converges to x in the weak topology, then, since $f \in X^*$ is continuous in the weak topology (by definition), $f(x_n) \to f(x)$. That is $x_n \overset{w}{\rightharpoonup} x$. Conversely, suppose $f(x_n) \to f(x)\ \forall\ f \in X^*$. Let U be a basic open set containing x. Then

$$U = x + \{y \in X : |f_i(y)| < \varepsilon_i,\ i = 1, \dots, m\}$$

for some m, $\varepsilon_i > 0$, and $f_i \in X^*$. Now there is some $N > 0$ such that

$$|f_i(x_n) - f_i(x)| = |f_i(x_n - x)| < \varepsilon_i$$

for all $n \geq N$, since $f_i(x_n) \to f_i(x)$, so $x_n = x + (x_n - x) \in U$. That is, x converges to x_n in the weak topology. Similar reasoning gives the result for weak-$*$ convergence. □

By Proposition 2.28, the weak and weak-$*$ topologies are Hausdorff. Obviously the weak topology on X is weaker than the *strong* or *norm* topology (for which more than just the linear functions are continuous).

On X^*, we have three topologies, the weak-$*$ topology (weakest for which the evaluation maps $\subset X^{**}$ are continuous), the weak topology (weakest for which X^{**} maps are continuous), and the strong or norm topology. The weak-$*$ topology is weaker than the weak topology, which is weaker than the strong topology. Of course, if X is reflexive, the weak-$*$ and weak topologies agree.

It is easier to obtain convergence in weaker topologies, as then there are fewer open sets to consider. In infinite dimensions, the unit ball is not a compact set. However, if we restrict the open sets in a cover to weakly open sets, we might hope to obtain compactness. This is in fact the case in X^*.

THEOREM 2.31 (Banach-Alaoglu Theorem). *Suppose X is a NLS with dual X^*, and B_1^* is the closed unit ball in X^* (i.e., $B_1^* = \{f \in X^* : \|f\| \leq 1\}$. Then B_1^* is compact in the weak-$*$ topology.*

By a scaling argument, we can immediately generalize the theorem to show that a closed unit ball of any radius $r > 0$ is weak-$*$ compact.

PROOF OF THE BANACH-ALAOGLU THEOREM. For each $x \in X$, let

$$B_x = \{\lambda \in \mathbb{F} : |\lambda| \leq \|x\|\} \ .$$

Each B_x is closed and bounded in $\mathbb{F}$, and so is compact. By Tychonoff's Theorem,

$$C = \mathop{\times}_{x \in X} B_x$$

is also compact. An element of C can be viewed as a function $g : X \to \mathbb{F}$ satisfying $|g(x)| \leq \|x\|$. In this way, B_1^* is the subset of C consisting of the linear functions. The product topology on C is the weakest one making all coordinate projection maps $g \mapsto g(x)$ continuous. As these maps are the evaluation maps, the inherited topology on B_1^* is precisely the weak-$*$ topology.

Since C is compact, we can complete the proof by showig that B_1^* is closed in C.

Since X^* is not a metric space when endowed with the weak-$*$ topology, we must consider an accumulation point g of B_1^*. Since every neighborhood of the form

$$U = g + \{h \in C : |h(x_i)| < \varepsilon_i, \ i = 1, \dots, m\}$$

intersects B_1^*, given $\varepsilon > 0$ and $x, y \in X$, we have an $f \in B_1^*$ such that

$$f = g + h$$

where

$$|h(x)| < \frac{\varepsilon}{3} \ , \quad |h(y)| < \frac{\varepsilon}{3} \ , \quad \text{and} \quad |h(x+y)| < \frac{\varepsilon}{3} \ .$$

Thus, since f is linear,

$$|g(x+y) - g(x) - g(y)| = |h(x+y) - h(x) - h(y)| \leq \varepsilon \ .$$

As ε is arbitrary, g is linear. Moreover,

$$|g(x)| = |f(x) - h(x)| \leq |f(x)| + \frac{\varepsilon}{3} \leq \|x\| + \frac{\varepsilon}{3} \ ,$$

so also $|g(x)| \le \|x\|$. That is, $g \in B_1^*$, so B_1^* is closed. $\square$

What does compactness say about sequences? If the space is metrizable (i.e., there is a metric that gives the same topology), a sequence in a compact space has a convergent subsequence (see Proposition 1.27).

THEOREM 2.32. *If X is a separable Banach space and $K \subset X^*$ is weak-$*$ compact, then K is metrizable in the weak-$*$ topology.*

PROOF. Separability means that we can find a dense subset $D = \{x_n\}_{n=1}^\infty \subset X$. The evaluation maps $E_n : X^* \to \mathbb{F}$, defined by $E_n(x^*) = x^*(x_n)$, are weak-$*$ continuous by definition. If $E_n(x^*) = E_n(y^*)$ for each n, then x^* and y^* are two continuous functions that agree on the dense set D, and so they must agree everywhere. That is, the set $\{E_n\}_{n=1}^\infty$ is a countable set of continuous functions that separates points on X^*.

Now let $C_n = \sup_{x^* \in K} |E_n(x^*)| < \infty$, since K is compact and E_n is continuous, and define $f_n = E_n/C_n$. Then $|f_n| \le 1$, and

$$d(x^*, y^*) = \sum_{n=1}^\infty 2^{-n} |f_n(x^*) - f_n(y^*)|$$

is a metric on K, since the f_n separate points.

We now have two topologies, the weak-$*$ open sets τ, and the open sets τ_d generated from the metric, which we must show coincide. That $\tau_d \subset \tau$ is easily seen, since any ball $B_r(y^*) = \{x^* \in K : d(x^*, y^*) < r\}$ is the inverse image of the open set $(-\infty, r)$ under the continuous function $d(\cdot, y^*)$ (for fixed y^*).

To show the opposite inclusion, $\tau \subset \tau_d$, let $A \in \tau$. Then $A^c \subset K$ is τ-closed, and thus τ-compact (Proposition 1.24). But $\tau_d \subset \tau$ implies that A^c is also τ_d-compact by definition, since any τ_d-open cover of A^c is also a τ-open cover, which has a finite subcover. Proposition 1.24 now implies that A^c is τ_d-closed, and thus $A \in \tau_d$. The proof is complete. $\square$

COROLLARY 2.33. *If X is a separable Banach space, $\{f_n\}_{n=1}^\infty \subset X^*$, and there is some $R > 0$ such that $\|f_n\| \le R$ for all n, then there is a subsequence $\{f_{n_i}\}_{i=1}^\infty$ that converges weak-$*$ in X^*.*

COROLLARY 2.34. *If Banach space X is separable and reflexive and $\{x_n\}_{n=1}^\infty \subset X$ is a bounded sequence, then there is a subsequence $\{x_n\}_{i=1}^\infty$ that converges weakly in X.*

COROLLARY 2.35 (Generalized Heine-Borel Theorem). *Suppose X is a Banach space with dual X^*, and $K \subset X^*$. Then K is weak-$*$ compact if and only if K is weak-$*$ closed and bounded.*

PROOF. Any (weak-$*$) closed and bounded set K is compact, as it sits in a large closed ball, which is compact. Conversely, if K is compact, it is closed. It must be bounded, for otherwise we can find a nonconvergent sequence in K (every weak-$*$ convergent sequence is bounded). $\square$

We close this section with an interesting result that relates weak and strong convergence.

THEOREM 2.36 (Banach-Saks). *Suppose that X is a NLS and $\{x_n\}_{n-1}^\infty$ is a sequence in X that converges weakly to $x \in X$. Then for every $n \ge 1$, there are constants $\alpha_j^n \ge 0$, $\sum_{j=1}^n \alpha_j^n = 1$, such that $y_n = \sum_{j=1}^n \alpha_j^n x_j$ converges strongly to x.*

That is, whenever $x_n \overset{w}{\rightharpoonup} x$, there is a sequence y_n of finite, convex, linear combinations of the x_n such that $y_n \to x$.

Proof. Let $z_n = x_n - x_1$ and $z = x - x_1$, so that $z_1 = 0$ is in the sequence and $z_n \overset{w}{\rightharpoonup} z$. Let

$$M = \left\{ \sum_{j=1}^{n} \alpha_j^n y_j : n \geq 1, \alpha_j^n \geq 0, \text{ and } \sum_{j=1}^{n} \alpha_j^n \leq 1 \right\} ,$$

which is convex. The conclusion of the theorem is that z is in $\bar{M}$, the (norm) closure of M. Suppose that this is not the case. Then we can apply the Separating Hyperplane Theorem 2.18 to the closed set $\bar{M}$ and the compact set $\{z\}$ to obtain a continuous linear functional f and a number γ such that $f(z_n) < \gamma$ but $f(z) > \gamma$. Thus $\limsup_{n\to\infty} f(z_n) \leq \gamma$, so $f(z_n) \not\to f(z)$, and we have a contradiction to $z_n \overset{w}{\rightharpoonup} z$ and must conclude that $z \in \bar{M}$ as required. □

Corollary 2.37. *Suppose that X is a NLS, and $S \subset X$ is convex. Then the weak and strong (norm) closures of S are identical.*

Proof. Let $\bar{S}^w$ denote the weak closure, and $\bar{S}$ the usual norm closure. The Banach-Saks Theorem implies that $\bar{S}^w \subset \bar{S}$, since S is convex. But trivially $\bar{S} \subset \bar{S}^w$. □

2.8. Conjugate or Dual of an Operator

Suppose X and Y are NLS's and $T \in B(X,Y)$. The operator T induces an operator

$$T^* : Y^* \to X^*$$

as follows. Let $g \in Y^*$ and define $T^* : X^* \to \mathbb{F}$ by the formula

$$(T^*g)(x) = g(Tx)$$

for $x \in X$. Then, $T^*g \in X^*$. For $T^*g = g \circ T$ is a composition of continuous linear maps,

$$\begin{tikzcd} X \arrow[r, "T"] \arrow[rr, bend right, "T^*g"'] & Y \arrow[r, "g"] & \mathbb{F} \end{tikzcd}$$

and so is itself continuous and linear. Moreover, if $g \in Y^*$, $x \in X$, then

$$\begin{aligned} |T^*g(x)| &= |g(Tx)| \leq \|g\|_{Y^*} \|Tx\|_Y \\ &\leq \|g\|_{Y^*} \|T\|_{B(X,Y)} \|x\|_X \\ &= \Big(\|g\|_{Y^*} \|T\|_{B(X,Y)} \Big) \|x\|_X \ . \end{aligned}$$

Hence, not only is T^*g bounded, but

$$\|T^*g\|_{X^*} \leq \|T\|_{B(X,Y)} \|g\|_{Y^*} \ . \tag{2.23}$$

Thus we have defined a correspondence $T^* : Y^* \to X^*$. In fact, T^* is itself a bounded linear map, which is to say $T^* \in B(Y^*, X^*)$. For linearity, we need to show that for $g, h \in Y^*$, $\lambda \in \mathbb{F}$,

$$\begin{aligned} T^*(g+h) &= T^*g + T^*h \ , \\ T^*(\lambda g) &= \lambda T^*g \ . \end{aligned} \tag{2.24}$$

Let $x \in X$ and evaluate both sides of these potential equalities at x, *viz.*

$$\begin{aligned} T^*(g+h)(x) &= (g+h)(Tx) = g(Tx) + h(Tx) \\ &= T^*(g)(x) + T^*h(x) \\ &= (T^*g(g) + T^*(h))(x) \end{aligned}$$

and

$$\begin{aligned} T^*(\lambda g)(x) &= (\lambda g)(Tx) = \lambda g(Tx) \\ &= \lambda T^* g(x) \ . \end{aligned}$$

As $x \in X$, was arbitrary, it follows that the formulas (2.24) are valid. Thus T^* is linear. The fact that T^* is bounded follows from (2.23), and, moreover,

$$\|T^*\|_{B(Y^*,X^*)} \leq \|T\|_{B(X,Y)} \ . \tag{2.25}$$

In fact, equality always holds in the last inequality. To see this, first note that if $T = 0$ is the zero operator, then $T^* = 0$ also and so their norms certainly agree. If $T \neq 0$, then $\|T\|_{B(X,Y)} > 0$. Let $\varepsilon > 0$ be given and let $x_0 \in X$, $\|x_0\|_X = 1$ be such that

$$\|Tx_0\|_Y \geq \|T\|_{B(X,Y)} - \varepsilon \ .$$

Let $g_0 \in Y^*$ be such that $\|g_0\|_{Y^*} = 1$ and

$$g_0(Tx_0) = \|Tx_0\| \ .$$

Such a g_0 exists by one of the corollaries of the Hahn-Banach Theorem. Then, it transpires that

$$\begin{aligned} \|T^*\|_{B(Y^*,X^*)} &\geq \|T^*g_0\|_{X^*} = \sup_{\|x\|_X=1} |T^*g_0(x)| \\ &\geq |T^*g_0(x_0)| = g_0(Tx_0) = \|Tx_0\|_Y \\ &\geq \|T\|_{B(X,Y)} - \varepsilon \ . \end{aligned}$$

In consequence of these ruminations, it is seen that

$$\|T^*\|_{B(Y^*,X^*)} \geq \|T\|_{B(X,Y)} - \varepsilon \ ,$$

and $\varepsilon > 0$ was arbitrary. Hence

$$\|T^*\|_{B(Y^*,X^*)} \geq \|T\|_{B(X,Y)}$$

and, along with (2.25), this establishes the result.

The $*$-mapping $T \longmapsto T^*$ is thus a norm preserving map

$$B(X,Y) \xrightarrow{*} B(Y^*,X^*) \ .$$

It has simple properties of its own, for example

$$\begin{aligned} &(\lambda T + \mu S)^* = \lambda T^* + \mu S^* \ , && \forall\ S,T \in B(X,Y) \ , \\ &(TS)^* = S^*T^* \ , && \forall\ S \in B(X,Y) \ , \ T \in B(Y,Z) \ , \\ &(I_X)^* = I_{X^*} \ , \end{aligned}$$

where I_X is the identity mapping of X to itself.

EXAMPLES. $X = \mathbb{R}^N$, $T : X \to X$ may be represented by an $N \times N$ matrix M_T in the standard basis, say. Then T^* also has a matrix representation in the dual basis and $M_{T^*} = M_T^t$ the transpose of M_T.

Here is a less elementary, but related example. Let $1 < p < \infty$ and, for $f \in L_p(0,1)$ and $x \in [0,1]$, set

$$Tf(x) = \int_0^1 K(x,y)f(y)\,dy \ ,$$

where K is, say, a bounded measurable function. It is easily determined that T is a bounded linear map of $L_p(0,1)$ into itself.

The dual space of $L_p(0,1)$ may be realized concretely as follows. If $\Lambda \in L_p^*(0,1)$, then there is a unique $g \in L_q(0,1)$, where

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

such that

$$\Lambda(f) = \int_0^1 f(x)g(x)\,dx \ .$$

Write $\Lambda = \Lambda_g$ in this case. What is T^*? If $\Lambda = \Lambda_g \in L_q(0,1)$ and $f \in L_p(0,1)$, then

$$\begin{aligned}(T^*\Lambda)(f) &= \Lambda(Tf) = \Lambda_g(Tf) \\ &= \int_0^1 g(x)Tf(x)\,dx = \int_0^1 g(x)\int_0^1 K(x,y)f(y)\,dy\,dx \\ &= \int_0^1 f(y)\int_0^1 K(x,y)g(x)\,dx\,dy \\ &= \int_0^1 f(y)T^*(\Lambda g)(y)\,dy \ .\end{aligned}$$

Thus, it is determined that

$$T^*(g)(y) = \int_0^1 K(x,y)g(x)\,dx \ .$$

LEMMA 2.38. *Let X, Y be NLS's and $T \in B(X,Y)$. Then $T^{**} : X^{**} \to Y^{**}$ is a bounded linear extension of T. If X is reflexive, then $T = T^{**}$.*

PROOF. Let $x \in X$ and $g \in Y^*$. Realize x as $[x] \in X^{**}$. Then, by definition,

$$\begin{aligned}(T^{**}[x])(g) &= [x](T^*g) = T^*g(x) \\ &= g(Tx) = [Tx](g) \ ,\end{aligned}$$

and so

$$T^{**}[x] = [Tx] \ .$$

Thus $T^{**}|_X = T$. If $X = X^{**}$, then this means $T = T^{**}$. □

LEMMA 2.39. *Let X be a Banach space, Y a NLS and $T \in B(X,Y)$. Then T has a bounded inverse defined on all of Y if and only if T^* has a bounded inverse defined on all of X^*. When either exists, then*

$$(T^{-1})^* = T^{*-1} \ .$$

PROOF. If $S = T^{-1} \in B(Y, X)$, then

$$S^*T^* = (TS)^* = (I_Y)^* = I_{Y^*} \ .$$

This shows that T^* is one-to-one. The other way around,

$$T^*S^* = (ST)^* = (T_X)^* = I_{X^*} \ ,$$

shows T^* is onto. Moreover, S^* is the inverse of T^*, and of course S^* is bounded since it is the dual of a bounded map.

Conversely, if $T^* \in B(Y^*, X^*)$ has a bounded inverse, then applying the preceding argument, we ascertain that $(T^{**})^{-1} \in B(Y^{**}, X^{**})$. But,

$$T^{**}\big|_X = T \ ,$$

so T must be one-to-one. Also, since T^{**} is onto, it is an open mapping and so $T^{**}(X)$ is closed in Y^{**} which is to say that $T(X)$ is closed in Y^{**}, and hence in Y. Suppose T is not onto. Let $y \in Y \smallsetminus T(X)$. By the Hahn-Banach Theorem, since $T(X)$ is closed, there is a $y^* \in Y^*$ such that

$$y^*\big|_{T(X)} = 0 \ , \quad \text{but} \quad y^*(y) \neq 0 \ .$$

But then, for all $x \in X$,

$$T^*y^*(x) = y^*(Tx) = 0 \ ,$$

whence $T^*y^* = 0$. But $y^* \neq 0$ in Y^*, and so T^* is not one-to-one, a contradiction. It is concluded that T is onto. □

2.9. Exercises

1. Suppose that X is a vector space.

   (a) If $A, B \subset X$ are convex, show that $A + B$ and $A \cap B$ are convex. What about $A \cup B$ and $A \setminus B$?

   (b) Show that $2A \subset A + A$. When is it true that $2A = A + A$?

2. Let (X, d) be a metric space.

   (a) Show that

   $$\rho(x, y) = \min(1, d(x, y))$$

   is also a metric.

   (b) Show that $U \subset X$ is open in (X, d) if and only if U is open in (X, ρ).

   (c) Repeat the above for

   $$\sigma(x, y) = \frac{d(x, y)}{1 + d(x, y)} \ .$$

3. Let X be a NLS, x_0 be a fixed vector in X, and $\alpha \neq 0$ a fixed scalar. Show that the mappings $x \mapsto x + x_0$ and $x \mapsto \alpha x$ are homeomorphisms of X onto itself.

4. Show that if X is a NLS, then X is homeomorphic to $B_r(0)$ for fixed r. [Hint: consider the mapping $x \mapsto \dfrac{xr}{1 + \|x\|}$.]

5. In $\mathbb{R}^d$, show that any two norms are equivalent. Hint: Consider the unit sphere, which is compact.

6. Let X and Y be NLS over the same field, both having the same finite dimension n. Then prove that X and Y are topologically isomorphic, where a topological isomorphism is defined to be a mapping that is simultaneously an isomorphism and a homeomorphism.

7. Show that $\big(C([a,b]), |\cdot|\big)$, the set of real-valued continuous functions in the interval $[a,b]$ with the sup-norm (L_∞-norm), is a Banach space.

8. If $f \in L_p(\Omega)$ show that

$$\|f\|_p = \sup \Big| \int_\Omega f\, g\, dx \Big| = \sup \int_\Omega |f\, g|\, dx$$

where the supremum is taken over all $g \in L_q(\Omega)$ such that $\|g\|_q \le 1$ and $1/p + 1/q = 1$, where $1 \le p, q \le \infty$.

9. Finite dimensional matrices.

(a) Let $M^{n\times m}$ be the set of matrices with real valued coefficients a_{ij}, for $1 \le i \le n$ and $1 \le j \le m$. For every $A \in M^{n\times m}$, define

$$\|A\| = \max_{x\in\mathbb{R}^m} \frac{|Ax|_{\mathbb{R}^n}}{|x|_{\mathbb{R}^m}} \ .$$

Show that $\big(M^{n\times m}, \|\cdot\|\big)$ is a NLS.

(b) Each $A \in M^{n\times n}$ defines a linear map of $\mathbb{R}^n$ into itself. Show that

$$\|A\| = \max_{|x|=|y|=1} y^T A x \ ,$$

where y^T is the transpose of y.

(c) Show that each $A \in M^{n\times n}$ is continuous.

10. Let X and Y be NLS and $T : X \to Y$ a linear mapping. We define the norm of T to be

$$\|T\| = \sup_{x\in X, x\neq 0} \frac{\|Tx\|_Y}{\|x\|_X} \ ,$$

which is finite. Show that

$$\|T\| = \sup_{\|x\|_X=1} \|Tx\|_Y = \sup_{\|x\|_X\le 1} \|Tx\|_Y = \inf\{M : \|Tx\|_Y \le M\|x\|_X \text{ for all } x \in X\} \ .$$

11. Let X be a vector space. We define the *convex hull* of $A \subset X$ to be

$$\operatorname{co}(A) = \{x \in X : x = \sum_{i=1}^n t_i y_i, \text{ where } t_i \in [0,1] \text{ and } y_i \in A\} \ .$$

(a) Prove that the convex hull is convex, and that it is the intersection of all convex subsets of X containing A.

(b) If X is a normed linear space, prove that the convex hull of an open set is open.

(c) If X is a normed linear space, is the convex hull of a closed set always closed?

(d) Prove that if X is a normed linear space, then the convex hull of a bounded set is bounded.

12. Prove that if X is a normed linear space and $B = B_1(0)$ is the unit ball, then X is infinite dimensional if and only if B contains an infinite collection of non-overlapping balls of diameter $1/2$.

13. Prove that a subset A of a metric space (X, d) is bounded if and only if every countable subset of A is bounded.

14. Consider $(\ell_p, |\cdot|_p)$.

    (a) Prove that ℓ_p is a Banach space for $1 \le p \le \infty$. Hint: Use that $\mathbb{R}$ is complete.

    (b) Show that $|\cdot|_p$ is *not* a norm for $0 < p < 1$. Hint: First show the result on $\mathbb{R}^2$.

15. If an infinite dimensional vector space X is also a NLS and contains a sequence $\{e_n\}_{n=1}^\infty$ with the property that for every $x \in X$ there is a unique sequence of scalars $\{\alpha_n\}_{n=1}^\infty$ such that
$$\|x - (\alpha_1 e_1 + \ldots + \alpha_n e_n\| \to 0 \quad \text{as } n \to \infty \ ,$$
then $\{e_n\}_{n=1}^\infty$ is called a *Schauder basis* for X, and we have the *expansion* of x
$$x = \sum_{n=1}^\infty \alpha_n e_n \ .$$

    (a) Find a Schauder basis for ℓ_p, $1 \le p < \infty$.

    (b) Show that if a NLS has a Schauder basis, then it is separable. [Remark: The converse is *not* true.]

16. Let Y be a subspace of a vector space X. The *coset* of an element $x \in X$ with respect to Y is denoted by $x + Y$ and is defined to be the set
$$x + Y = \{z \in X : z = x + y \text{ for some } y \in Y\} \ .$$
Show that the distinct cosets form a partition of X. Show that under algebraic operations defined by
$$(x_1 + Y) + (x_2 + Y) = (x_1 + x_2) + Y \quad \text{and} \quad \lambda(x + Y) = \lambda x + Y \ ,$$
for any $x_1, x_2, x \in X$ and λ in the field, these cosets form a vector space. This space is called the *quotient space* of X by (or *modulo* Y, and it is denoted X/Y.

17. Let Y be a closed subspace of a NLS X. Show that a norm on the quotient space X/Y is given for $\hat{x} \in X/Y$ by
$$\|\hat{x}\|_{X/Y} = \inf_{x \in \hat{x}} \|x\|_X \ .$$

18. If X and Y are NLS, then the product space $X \times Y$ is also a NLS with any of the norms
$$\|(x, y)\|_{X \times Y} = \max(\|x\|_X, \|y\|_Y)$$
or, for any $1 \le p < \infty$,
$$\|(x, y)\|_{X \times Y} = (\|x\|_X^p + \|y\|_Y^p)^{1/p} \ .$$
Why are these norms equivalent?

19. If X and Y are Banach spaces, prove that $X \times Y$ is a Banach space.

20. Let $T : C([0, 1]) \to C([0, 1])$ be defined by
$$y(t) = \int_0^t x(\tau)\, d\tau \ .$$

Find the range $R(T)$ of T, and show that T is invertible on its range, $T^{-1} : R(T) \to C([0,1])$. Is T^{-1} linear and bounded?

21. Show that on $C([a,b])$, for any $y \in C([a,b])$ and scalars α and β, the functionals
$$f_1(x) = \int_a^b x(t)\, y(t)\, dt \quad \text{and} \quad f_2(x) = \alpha x(a) + \beta x(b)$$
are linear and bounded.

22. Find the norm of the linear functional f defined on $C([-1,1])$ by
$$f(x) = \int_{-1}^{0} x(t)\, dt - \int_0^1 x(t)\, dt\ .$$

23. Recall that $f = \{\{x_n\}_{n=1}^\infty : \text{only finitely many } x_n \neq 0\}$ is a NLS with the sup-norm $|\{x_n\}| = \sup_n |x_n|$. Let $T : f \to f$ be defined by
$$T(\{x_n\}_{n=1}^\infty) = \{n x_n\}_{n=1}^\infty\ .$$
Show that T is linear but *not* continuous (i.e., not bounded).

24. The space $C^1([a,b])$ is the NLS of all continuously differentiable functions defined on $[a,b]$ with the norm
$$\|x\| = \sup_{t\in[a,b]} |x(t)| + \sup_{t\in[a,b]} |x'(t)|\ .$$
(a) Show that $\|\cdot\|$ is indeed a norm.
(b) Show that $f(x) = x'\big((a+b)/2\big)$ defines a continuous linear functional on $C^1([a,b])$.
(c) Show that f defined above is *not* bounded on the subspace of $C([a,b])$ consisting of all continuously differentiable functions with the norm inherited from $C([a,b])$.

25. Suppose X is a vector space. The *algebraic dual* of X is the set of all linear functionals on X, and is also a vector space. Suppose also that X is a NLS. Show that X has finite dimension if and only if the algebraic dual and the dual space X^* coincide.

26. Let X be a NLS and M a nonempty subset. The *annihilator* M^a of M is defined to be the set of all bounded linear functionals $f \in X^*$ such that f restricted to M is zero. Show that M^a is a closed subspace of X^*. What are X^a and $\{0\}^a$?

27. Define the operator T by the formula
$$T(f)(x) = \int_a^b K(x,y)\, f(y)\, dy\ .$$
Suppose that $K \in L_q([a,b]\times[a,b])$, where q lies in the range $1 \le q \le \infty$. Determine the values of p for which T is necessarily a bounded linear operator from $L_p(a,b)$ to $L_q(a,b)$. In particular, if a and b are both finite, show that $K \in L_\infty([a,b]\times[a,b])$ implies T to be bounded on all the L_p-spaces.

28. Let $U = B_r(0) = \{x : \|x\| < r\}$ be an open ball about 0 in a real normed linear space, and let $y \notin \bar{U}$. Show that there is a bounded linear functional f that separates U from y. (That is, U and y lie in opposite half spaces determined by f, which is to say there is an α such that U lies in $\{x : f(x) < \alpha\}$ and $f(y) > \alpha$.)

29. Prove that $L_2([0,1])$ is of the first category in $L_1([0,1])$. (Recall that a set is of first category if it is a countable union of nowhere dense sets, and that a set is nowhere dense if its closure

has an empty interior.) Hint: Show that $A_k = \{f : \|f\|_{L_2} \le k\}$ is closed in L_1 but has empty interior.

30. If a Banach space X is reflexive, show that X^* is also reflexive. ($*$) Is the converse true? Give a proof or a counterexample.

31. Let $y = (y_1, y_2, y_3, ...) \in \mathbb{C}^\infty$ be a vector of complex numbers such that $\sum_{i=1}^\infty y_i x_i$ converges for every $x = (x_1, x_2, x_3, ...) \in \mathbb{C}_0$, where $\mathbb{C}_0 = \{x \in \mathbb{C}^\infty : x_i \to 0 \text{ as } i \to \infty\}$. Prove that

$$\sum_{i=1}^\infty |y_i| < \infty \ .$$

32. Let X and Y be normed linear spaces, $T \in B(X,Y)$, and $\{x_n\}_{n=1}^\infty \subset X$. If $x_n \overset{w}{\rightharpoonup} x$, prove that $Tx_n \overset{w}{\rightharpoonup} Tx$ in Y. Thus a bounded linear operator is weakly sequentially continuous. Is a weakly sequentially continuous linear operator necessarily bounded?

33. Suppose that X is a Banach space, M and N are linear subspaces, and that $X = M \oplus N$, which means that

$$X = M + N = \{m + n : m \in M, n \in N\}$$

and $M \cap N$ is the trivial linear subspace consisting only of the zero element. Let P denote the projection of X onto M. That is, if $x = m + n$, then

$$P(x) = m$$

Show that P is well defined and linear. Prove that P is bounded if and only if both M and N are closed.

34. Let X be a Banach space, Y a NLS, and $T_n \in B(X,Y)$ such that $\{T_n x\}_{n=1}^\infty$ is a Cauchy sequence in Y. Show that $\{\|T_n\|\}_{n=1}^\infty$ is bounded. If, in addition, Y is a Banach space, show that if we define T by $T_n x \to Tx$, then $T \in B(X,Y)$.

35. Let X be the normed space of sequences of complex numbers $x = \{x_i\}_{i=1}^\infty$ with only finitely many nonzero terms and norm defined by $\|x\| = \sup_i |x_i|$. Let $T : X \to X$ be defined by

$$y = Tx = \{x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, ...\} \ .$$

Show that T is a bounded linear map, but that T^{-1} is unbounded. Why does this *not* contradict the Open Mapping Theorem?

36. Give an example of a function that is closed but *not* continuous.

37. For each $\alpha \in \mathbb{R}$, let E_α be the set of all continuous functions f on $[-1,1]$ such that $f(0) = \alpha$. Show that the E_α are convex, and that each is dense in $L_2([-1,1])$.

38. Suppose that X, Y, and Z are Banach spaces and that $T : X \times Y \to Z$ is bilinear and continuous. Prove that there is a constant $M < \infty$ such that

$$\|T(x,y)\| \le M\|x\|\,\|y\| \quad \text{for all } x \in X, y \in Y \ .$$

Is completeness needed here?

39. Prove that a bilinear map is continuous if it is continuous at the origin $(0,0)$.

40. Consider $X = C([a,b])$, the continuous functions defined on $[a,b]$ with the maximum norm. Let $\{f_n\}_{n=1}^\infty$ be a sequence in X and suppose that $f_n \xrightarrow{w} f$. Prove that $\{f_n\}_{n=1}^\infty$ is pointwise convergent. That is,
$$f_n(x) \to f(x) \quad \text{for all } x \in [a,b] \ .$$
Prove that a weakly convergent sequence in $C^1([a,b])$ is convergent in $C([a,b])$. $(*)$ Is this still true when $[a,b]$ is replaced by $\mathbb{R}$?

41. Let X be a normed linear space and Y a closed subspace. Show that Y is weakly sequentially closed.

42. Let X be a normed linear space. We say that a sequence $\{x_n\}_{n=1}^\infty \subset X$ is *weakly Cauchy* if $\{Tx_n\}_{n=1}^\infty$ is Cauchy for all $T \in X^*$, and we say that X is *weakly complete* if each weak Cauchy sequence converges weakly. If X is reflexive, prove that X is weakly complete.

43. Show that every finite dimensional vector space is reflexive.

44. Show that $C([0,1])$ is not reflexive.

45. If X and Y are Banach spaces, show that $E \subset B(X,Y)$ is equicontinuous if, and only if, there is an $M < \infty$ such that $\|T\| \leq M$ for all $T \in E$.

46. Let X be a Banach space and $T \in X^* = B(X,\mathbb{F})$. Identify the range of $T^* \in B(\mathbb{F}, X^*)$.

47. Let X be a Banach space, $S, T \in B(X,X)$, and I be the identity map.

    (a) Show by example that $ST = I$ does not imply $TS = I$.

    (b) If T is compact, show that $S(I-T) = I$ if, and only if, $(I-T)S = I$.

    (c) If $S = (I-T)^{-1}$ exists for some T compact, show that $I - S$ is compact.

48. Let $1 \leq p < \infty$ and define, for each $r \in \mathbb{R}^d$, $T_r : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ by
$$T_r(f)(x) = f(x+r) \ .$$

    (a) Verify that $T_r(f) \in L_p(\mathbb{R}^d)$ and that T_r is bounded and linear. What is the norm of T_r?

    (b) Show that as $r \to s$, $\|T_r f - T_s f\|_{L_p} \to 0$. Hint: Use that the set of continuous functions with compact support are dense in $L_p(\mathbb{R}^d)$ for $p < \infty$.

CHAPTER 3

Hilbert Spaces

The norm of a normed linear space gives a notion of absolute size for the elements of the space. While this has generated an extremely interesting and useful structure, often one would like more geometric information about the elements. In this chapter we add to the NLS structure a notion of "angle" between elements and, in particular, a notion of orthogonality through a device known as an inner-product.

3.1. Basic properties of inner-products

DEFINITION. An *inner-product* on a vector space H is a map $(\cdot,\cdot) : H \times H \to \mathbb{F}$ satisfying the following properties.

(a) The map $(\cdot,\cdot)$ is linear in its first argument; that is, for $\alpha, \beta \in \mathbb{F}$ and $x, y, z \in H$,

$$(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z) \ .$$

(b) The map $(\cdot,\cdot)$ is *conjugate symmetric* (*symmetric* if $\mathbb{F} = \mathbb{R}$), meaning that for $x, y \in H$,

$$(x, y) = \overline{(y, x)} \ .$$

(c) For any $x \in H$, $(x, x) \geq 0$; moreover, $(x, x) = 0$ if and only if $x = 0$.

If H has such an inner-product, then H is called an *inner-product space* (IPS) or a *pre-Hilbert space*. Any map satisfying (a) and (b) is said to be *sesquilinear* (*bilinear* if $\mathbb{F} = \mathbb{R}$).

PROPOSITION 3.1. *If $(\cdot,\cdot)$ is sesquilinear on H, then for $\alpha, \beta \in \mathbb{F}$ and $x, y, z \in \mathbb{F}$,*

$$(x, \alpha y + \beta z) = \overline{\alpha}(x, y) + \overline{\beta}(x, z) \ .$$

That is, $(\cdot,\cdot)$ is conjugate linear *in its second argument.*

EXAMPLES. (a) $\mathbb{C}^d$ (or $\mathbb{R}^d$) is an IPS with the inner-product

$$(x, y) = x \cdot \bar{y} = \sum_{i=1}^{d} x_i \bar{y}_i$$

for any $x, y \in \mathbb{C}^d$.

(b) Similarly ℓ_2 is an IPS with

$$(x, y) = \sum_{i=1}^{\infty} x_i \bar{y}_i \ .$$

(c) For any domain $\Omega \subset \mathbb{R}^d$, $L_2(\Omega)$ has inner-product

$$(f, g) = \int_{\Omega} f(x)\, \overline{g(x)}\, dx \ .$$

Definition. If $(H, (\cdot,\cdot))$ is an IPS, we define the map $\|\cdot\| : H \to \mathbb{R}$ by
$$\|x\| = (x,x)^{1/2}$$
for any $x \in H$. This map is called the *induced norm*.

Lemma 3.2 (Cauchy-Schwarz Inequality). *If $(H, (\cdot,\cdot))$ is an IPS with induced norm $\|\cdot\|$, then for any $x, y \in H$,*
$$|(x,y)| \le \|x\|\,\|y\| \ ,$$
with equality holding if and only if x or y is a multiple of the other.

Proof. If $y = 0$, there is nothing to prove, so assume $y \neq 0$. Then for $x, y \in H$ and $\lambda \in \mathbb{F}$,
$$\begin{aligned} 0 \le \|x - \lambda y\|^2 &= (x - \lambda y,\, x - \lambda y) \\ &= (x,x) - \overline{\lambda}(x,y) - \lambda(y,x) + |\lambda|^2 (y,y) \\ &= \|x\|^2 - (\,\overline{\lambda(y,x)} + \lambda(y,x)) + |\lambda|^2 \|y\|^2 \\ &= \|x\|^2 - 2\ \mathrm{Real}(\lambda(y,x)) + |\lambda|^2\|y\|^2 \ . \end{aligned}$$
Let
$$\lambda = \frac{(x,y)}{\|y\|^2} \ .$$
Then
$$0 \le \|x\|^2 - 2\ \mathrm{Real}\ \frac{(x,y)(y,x)}{\|y\|^2} + \frac{|(x,y)|^2}{\|y\|^4}\,\|y\|^2 = \|x\|^2 - \frac{|(x,y)|^2}{\|y\|^2} \ ,$$
since $(x,y)(y,x) = |(x,y)|^2$ is real. A rearrangement gives the result, with equality only if $x - \lambda y = 0$. $\square$

Corollary 3.3. *The induced norm is indeed a norm, and thus an IPS H is a NLS.*

Proof. For $\alpha \in \mathbb{F}$ and $x \in H$, $\|x\| \ge 0$, $\|\alpha x\| = (\alpha x, \alpha x)^{1/2} = |\alpha|(x,x)^{1/2} = |\alpha|\,\|x\|$ and $\|x\| = 0$ if and only if $(x,x) = 0$ if and only if $x = 0$.

It remains only to demonstrate the triangle inequality. For $x, y \in H$,
$$\begin{aligned} \|x+y\|^2 &= (x+y, x+y) \\ &= \|x\|^2 + 2\ \mathrm{Real}(x,y) + \|y\|^2 \\ &\le \|x\|^2 + 2|(x,y)| + \|y\|^2 \\ &\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 \\ &= (\|x\| + \|y\|)^2 \ . \end{aligned}$$
$\square$

Note that the Cauchy-Schwarz inequality gives a notion of angle, as we may define the angle θ between x and y from
$$\cos\theta = \frac{|(x,y)|}{\|x\|\,\|y\|} \le 1 \ .$$
However, generally we consider only the case where $\theta = \pi/2$.

Definition. If $(H, (\cdot,\cdot))$ is an IPS, $x, y \in H$, and $(x,y) = 0$, then we say that x and y are *orthogonal*, and denote this fact as $x \perp y$.

PROPOSITION 3.4 (Parallelogram Law). *If $x, y \in H$, an IPS, then*

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

PROOF. Exercise. $\square$

The parallelogram law can be used to show that not all norms come from an inner-product, as there are norms that violate the law. The law expresses the geometry of a parallelogram in $\mathbb{R}^2$, generalized to an arbitrary IPS.

LEMMA 3.5. *If $(H, \langle\cdot,\cdot\rangle)$ is an IPS, then $\langle\cdot,\cdot\rangle : H \times H \to \mathbb{F}$ is continuous.*

PROOF. Since $H \times H$ is a metric space, it is enough to show sequential continuity. So suppose that $(x_n, y_n) \to (x, y)$ in $H \times H$; that is, both $x_n \to x$ and $y_n \to y$. Then

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

Since $x_n \to x$, $\|x_n\|$ is bounded. Thus $|\langle x_n, y_n\rangle - \langle x, y\rangle|$ can be made as small as desired by taking n sufficiently large. $\square$

COROLLARY 3.6. *If $\lambda_n \to \lambda$ and $\mu_n \to \mu$ in $\mathbb{F}$ and $x_n \to x$ and $y_n \to y$ in H, then*

$$\langle \lambda_n x_n, \mu_n y_n\rangle \to \langle \lambda x, \mu y\rangle \ .$$

PROOF. Just note that $\lambda_n x_n \to \lambda x$ and $\mu_n y_n \to \mu y$. $\square$

DEFINITION. A complete IPS H is called a *Hilbert* space.

Hilbert spaces are thus Banach spaces.

3.2. Best approximation and orthogonal projections

The following is an important geometric relation in an IPS.

THEOREM 3.7 (Best approximation). *Suppose $(H, (\cdot,\cdot))$ is an IPS and $M \subset H$ is nonempty, convex, and complete (i.e., closed if H is Hilbert). If $x \in H$, then there is a unique $y = y(x) \in M$ such that*

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

We call y the best approximation *of or* closest point *to x from M.*

PROOF. Let

$$\delta = \inf_{Z \in M} \|x - z\| \ .$$

If $\delta = 0$, we must take $x = y$. That $y = x$ is in M follows from completeness, since given any integer $n \ge 1$, there is some $z_n \in M$ such that $\|x - z_n\| = 1/n$, so $z_n \to x \in M$.

Suppose $\delta > 0$. Then $x \notin M$ and so there is a sequence $\{y_n\}_{n=1}^\infty \subset M$ such that as $n \to \infty$,

$$\|x - y_n\| \equiv \delta_n \to \delta \ .$$

We claim that $\{y_n\}$ is Cauchy. By the parallelogram law,

$$\begin{aligned}\|y_n - y_m\|^2 &= \|(y_n - x) + (x - y_m)\|^2 \\ &= 2\left(\|y_n - x\|^2 + \|x - y_m\|^2\right) - \|y_n + y_m - 2x\|^2 \\ &= 2(\delta_n^2 + \delta_m^2) - 4\left\|\frac{y_n + y_m}{2} - x\right\|^2 \\ &\le 2(\delta_n^2 + \delta_m^2) - 4\delta^2 \ ,\end{aligned}$$

since by convexity $(y_n + y_m)/2 \in M$. Thus as $n, m \to \infty$, $\|y_n - y_m\| \to 0$. By completeness, $y_n \to y$ for some $y \in M$. Since $\|\cdot\|$ is continuous, $\|x - y\| = \delta$.

To see that y is unique, suppose that for some $z \in M$, $\|x - z\| = \delta$. Then the parallelogram law again shows

$$\begin{aligned}\|y - z\|^2 &= \|(y - x) + (x - z)\|^2 \\ &= 2\left(\|y - x\|^2 + \|x - z\|^2\right) - \|y + z - 2x\|^2 \\ &= 4\delta^2 - 4\left\|\frac{y + z}{2} - x\right\|^2 \\ &\le 4\delta^2 - 4\delta^2 = 0 \ .\end{aligned}$$

Thus $y = z$. □

Corollary 3.8. *Suppose $(H, (\cdot,\cdot))$ is an IPS and M is a complete linear subspace. If $x \in H$ and $y \in M$ is the best approximation to x in M, then*

$$x - y \perp M \ .$$

Proof. Let $m \in M$, $m \neq 0$. For any $\lambda \in \mathbb{F}$, by best approximation,

$$\|x - y\|^2 \le \|x - y + \lambda m\|^2 = \|x - y\|^2 + \bar\lambda(x - y, m) + \lambda(m, x - y) + |\lambda|^2\|m\|^2$$

With $\lambda = -(x - y, m)/\|m\|^2$, we have

$$0 \le -\bar\lambda\lambda\|m\|^2 - \lambda\bar\lambda\|m\|^2 + |\lambda|^2\|m\|^2 = -|\lambda|^2\|m\|^2 \ ,$$

so $\lambda = 0$, which means

$$(x - y, m) = 0$$

for any $m \in M$. That is, $x - y \perp M$. □

Definition. Given an IPS H and $M \subset H$,

$$M^\perp = \{x \in H : (x, m) = 0 \ \forall \ m \in M\} \ .$$

The space $M^\perp$ is referred to as "M-perp."

Proposition 3.9. *Suppose H is an IPS and $M \subset H$. Then $M^\perp$ is a linear subspace of H, $M \perp M^\perp$, and $M \cap M^\perp$ is either $\{0\}$ or $\emptyset$.*

Theorem 3.10. *Suppose $(H, (\cdot,\cdot))$ is an IPS and $M \subset H$ is a complete linear subspace. Then there exist two unique bounded linear surjective mappings*

$$P : H \to M \quad \text{and} \quad P^\perp : H \to M^\perp$$

defined by (a) *and* (b) *below and having the properties* (c)–(g) *for any $x \in H$*

(a) $\|x - Px\| = \inf_{y \in M} \|x - y\|$ *(i.e., Px is the best approximation to x in M),*
(b) $x = Px + P^\perp x$ *(i.e., $P^\perp = I - P$),*

(c) $\|x\|^2 = \|Px\|^2 + \|P^\perp x\|^2$,
(d) $x \in M$ *if and only if* $P^\perp x = 0$ *(i.e.,* $x = Px$*)*,
(e) $x \in M^\perp$ *if and only if* $Px = 0$ *(i.e.,* $x = P^\perp x$*)*,
(f) $\|P\| = 1$ *unless* $M = \{0\}$*, and* $\|P^\perp\| = 1$ *unless* $M = H$*,*
(g) $PP^\perp = P^\perp P = 0$*,* $P^2 = P$*, and* $(P^\perp)^2 = P^\perp$ *(i.e.,* P *and* $P^\perp$ *are* orthogonal projection *operators).*

Note that (c) is the Pythagorean theorem in an IPS, since $Px \perp P^\perp x$ and (b) holds. We call P and $P^\perp$ the *orthogonal projections* of H onto M and $M^\perp$, respectively.

PROOF. By the best approximation theorem, (a) defines P uniquely, and then (b) defines $P^\perp : H \to H$ uniquely. But if $x \in H$, then for $m \in M$,

$$(P^\perp x, m) = (x - Px, m) = 0$$

by Corollary 3.8, so the range of $P^\perp$ is $M^\perp$.

To see that P and $P^\perp$ are linear, let $\alpha, \beta \in \mathbb{F}$ and $x, y \in H$. Then by (b),

$$\alpha x + \beta y = P(\alpha x + \beta y) + P^\perp(\alpha x + \beta y) \ ,$$

and

$$\begin{aligned} \alpha x + \beta y &= \alpha(Px + P^\perp x) + \beta(Py + P^\perp y) \\ &= \alpha Px + \beta Py + \alpha P^\perp x + \beta P^\perp y \ . \end{aligned}$$

Thus

$$\alpha Px + \beta Py - P(\alpha x + \beta y) = P^\perp(\alpha x + \beta y) - \alpha P^\perp x - \beta P^\perp y \ .$$

Since M and $M^\perp$ are vector spaces, the left side above is in M and the right side is in $M^\perp$. So both sides are in $M \cap M^\perp = \{0\}$, and so

$$\begin{aligned} P(\alpha x + \beta y) &= \alpha Px + \beta Py \ , \\ P^\perp(\alpha x + \beta y) &= \alpha P^\perp x + \beta P^\perp y \ ; \end{aligned}$$

that is, P and $P^\perp$ are linear.

From the proof of the best approximation theorem, we saw that if $x \in M$, then $Px = x$; thus, P is surjective. Also, $x = Px$ implies $x = Px \in M$, so (d) follows.

If $x \in M^\perp$, then since $x = Px + P^\perp x$,

$$x - P^\perp x = Px \in M \cap M^\perp = \{0\} \ ,$$

so $x \in P^\perp x$, $P^\perp$ is surjective, and (e) follows.

If $x \in H$ then (e) and (d) imply that $PP^\perp x = 0$ since $P^\perp x \in M^\perp$ and $P^\perp Px = 0$ since $Px \in M$, so $0 = PP^\perp = P(I - P) = P - P^2$ and $0 = P^\perp P = P^\perp(I - P^\perp) = P^\perp - (P^\perp)^2$. That is, (g) follows.

We obtain (c) by direct computation,

$$\begin{aligned} \|x\|^2 = \|Px + P^\perp x\|^2 &= (Px + P^\perp x, Px + P^\perp x) \\ &= \|Px\|^2 + (Px, P^\perp x) + (P^\perp x, Px) + \|P^\perp x\|^2 \ . \end{aligned}$$

The two cross terms on the left vanish since $M \perp M^\perp$.

Finally, (c) implies that

$$\|Px\|^2 = \|x\|^2 - \|P^\perp x\| \leq \|x\|^2 \ ,$$

so $\|P\| \leq 1$. But if $M \neq \{0\}$, there exists $x \in M \setminus \{0\}$ for which $\|Px\| = \|x\|$. Thus $\|P\| = 1$. Similarly remarks apply to $P^\perp$. We conclude that P and $P^\perp$ are bounded and (f) holds. $\square$

Corollary 3.11. *If $(H, (\cdot, \cdot))$ is a Hilbert space and $M \subset H$ is a closed linear subspace, then $P^\perp$ is best approximation of H in $M^\perp$.*

Proof. We have the unique operators P_M and $(P_M)^\perp$ from the theorem. Now it is easy to verify that $M^\perp$ is closed, since the inner-product is continuous, so we can apply the theorem also to $M^\perp$ to obtain the unique operators $P_{M^\perp}$ and $(P_{M^\perp})^\perp$. It is not difficult to conclude that $P_{M^\perp} = (P_M)^\perp$, which is best approximation of H in $M^\perp$. $\square$

3.3. The dual space

We turn now to a discussion of the dual H^* of a Hilbert space $(H, (\cdot, \cdot))$. We first observe that if $y \in H$, then the functional L_y defined by

$$L_y(x) = (x, y)$$

is linear in x and bounded by the Cauchy-Schwarz inequality. In fact,

$$|L_y(x)| \leq \|y\| \, \|x\| \ ,$$

so $\|L_y\| \leq \|y\|$. But $|L_y(y/\|y\|)| = \|y\|$, so in fact

$$\|L_y\| = \|y\| \ .$$

We conclude that $L_y \in H^*$, and, as y is arbitrary,

$$\{L_y\}_{y \in H} \subset H^* \ .$$

We have represented certain members of H^* as L_y maps; in fact, as we will see, every member of H^* can be so represented. Thus by identifying L_y with y, we see that in some sense H is its own dual.

Theorem 3.12 (Riesz Representation Theorem). *Let $(H, (\cdot, \cdot))$ be a Hilbert space and $L \in H^*$. Then there is a unique $y \in H$ such that*

$$Lx = (x, y) \quad \forall \ x \in H \ .$$

Moreover, $\|L\|_{H^*} = \|y\|_H$.

Proof. If $L \equiv 0$ (i.e., $Lx = 0 \ \forall \ x \in H$), then take $y = 0$. Uniqueness is clear, since if $Lx = (x, z)$, then

$$0 = Lz = (z, z) = \|z\|^2$$

implies $z = 0$.

Suppose then that $L \not\equiv 0$. Let

$$M = N(L) \equiv \ker(L) \equiv \{x \in H : Lx = 0\} \ .$$

As M is the inverse image of the closed set $\{0\}$ under L, M is closed. Easily M is a vector space, so M is a closed (i.e., complete) linear subspace of H.

Since $L \not\equiv 0$, $M \neq H$ and $M^\perp \neq \{0\}$ by Theorem 3.10. Let $z \in M^\perp \setminus \{0\}$, normalized so $\|z\| = 1$. For $x \in H$, let

$$u = (Lx)z - (Lz)x \ ,$$

so

$$Lu = (Lx)(Lz) - (Lz)(Lx) = 0 \ .$$

Thus $u \in M$ and so $u \perp z$. That is,

$$0 = (u, z) = \Big((Lx)z - (Lz)x, z\Big) = Lx(z, z) - Lz(x, z) = Lx - Lz(x, z) \ ,$$

or

$$Lx = Lz(x, z) = (x, (\overline{Lz})z) \ .$$

Uniqueness is trivial, for if

$$Lx = (x, y_1) = (x, y_2) \quad \forall\, x \in H \ ,$$

then

$$(x, y_1 - y_2) = 0 \quad \forall\, x \in H \ .$$

Substitute $x = y_1 - y_2$ to conclude $y_1 = y_2$. Finally, we already saw that $\|L\| = \|L_y\| = \|y\|$. ☐

We define a map $R : H \to H^*$, called the *Riesz map*, by

$$Rx = L_x \ \forall\, x \in H \ .$$

The Riesz Representation Theorem says that R is one-to-one and onto. Thus we identify H with its dual precisely through R: Given $x \in H$ there is a unique $Rx = L_x \in H^*$, and conversely given $L \in H^*$, there is a unique $x = R^{-1}L \in H$ such that $L = L_x$. While R is not linear when $\mathbb{F} = \mathbb{C}$, it is conjugate linear:

$$R(x + y) = Rx + Ry \quad \forall\, x, y \in H \ ,$$
$$R(\lambda x) = \overline{\lambda} Rx \quad \forall\, x \in H,\ \lambda \in \mathbb{F} \ .$$

3.4. Orthonormal subsets

In finite dimensions, a vector space is isomorphic to $\mathbb{R}^d$ for some $d < \infty$, which can be described by an orthogonal basis. Similar results hold for infinite dimensional Hilbert spaces.

Definition. Suppose H is an IPS and $\mathcal{I}$ is some index set. A set $A = \{x_\alpha\}_{\alpha \in \mathcal{I}} \subset H$ is said to be *orthogonal* if $x_\alpha \neq 0 \ \forall\, \alpha \in \mathcal{I}$ and

$$x_\alpha \perp x_\beta \qquad \text{(i.e., } (x_\alpha, x_\beta) = 0)$$

for all $\alpha, \beta \in \mathcal{I}$, $\alpha \neq \beta$. Furthermore, if also $\|x_\alpha\| = 1 \ \forall\, \alpha \in \mathcal{I}$, then A is *orthonormal* (ON).

Definition. If $A \subset H$, a Hilbert space, then A is linearly independent if every finite subset of A is linearly independent. That is, every collection $\{x_i\}_{i=1}^n \subset A$ must satisfy the property that if there are scalars $c_i \in \mathbb{F}$ with

$$\sum_{i=1}^n c_i x_i = 0 \tag{3.1}$$

then necessarily $c_i = 0 \ \forall\, i$.

Proposition 3.13. *If a subset A of a Hilbert space H is orthogonal, then A is linearly independent.*

Proof. If $\{x_i\}_{i=1}^n \subset A$ and $c_i \in \mathbb{F}$ satisfy (3.1), then for $1 \leq j \leq n$,

$$0 = \left(\sum_{i=1}^n c_i x_i, x_j\right) = \sum_{i=1}^n c_i (x_i, x_j) = c_j \|x_j\|^2 \ .$$

As $x_j \neq 0$, necessarily each $c_j = 0$. ☐

Let $\{x_1, \dots, x_n\}$ be linearly independent in a Hilbert space H, and

$$M = \operatorname{span}\{x_1, \dots, x_n\} ,$$

which is closed in H as it is finite dimensional. We compute the orthogonal projection of $x \in H$ onto M. That is, we want $c_1, \dots, c_n \in \mathbb{F}$ such that $P_M x = \sum_{j=1}^n c_j x_j$ and $P_M x - x \perp M$. That is, for every $1 \le i \le n$,

$$(P_M x, x_i) = (x, x_i) .$$

Now

$$(P_M x, x_i) = \sum_{j=1}^n c_j (x_j, x_i) ,$$

so with

$$a_{ij} = (x_i, x_j) \quad \text{and} \quad b_i = (x, x_i)$$

we have that the $n \times n$ matrix $A = (a_{ij})$ and n-vectors $b = (b_i)$ and $c = (c_j)$ satisfy

$$Ac = b .$$

We already know that a unique solution c exists, so A is invertible and the solution c can be found, giving $P_M x$.

THEOREM 3.14. *Suppose H is a Hilbert space and $\{u_1, \dots, u_n\} \subset H$ is* ON. *Let $x \in H$. Then the orthogonal projection of x onto $M = \operatorname{span}\{u_1, \dots, u_n\}$ is given by*

$$P_M x = \sum_{i=1}^n (x, u_i) u_i .$$

Moreover,

$$\sum_{i=1}^n |(x, u_i)|^2 \le \|x\|^2 .$$

PROOF. In this case, the matrix $A = ((u_i, u_j)) = I$, so our coefficients c are the values $b = ((x, u_i))$. The final remark follows from the fact that $\|P_M x\| \le \|x\|$ and the calculation

$$\|P_M x\|^2 = \sum_{i=1}^n |(x, u_i)|^2 ,$$

left to the reader. □

We extend this result to larger ON sets. To do so, we need to note a few facts about infinite series. Let $\mathcal{I}$ be any index set (possibly uncountable!), and $\{x_\alpha\}_{\alpha \in \mathcal{I}}$ a series of nonnegative real numbers. We define

$$\sum_{\alpha \in \mathcal{I}} x_\alpha = \sup_{\substack{\mathcal{J} \subset \mathcal{I} \\ \mathcal{J} \text{ finite}}} \sum_{\alpha \in \mathcal{J}} x_\alpha .$$

If $\mathcal{I} = \mathbb{N} = \{0, 1, 2, \dots\}$ is countable, this agrees with the usual definition

$$\sum_{\alpha=0}^\infty x_\alpha = \lim_{n \to \infty} \sum_{\alpha=0}^n x_\alpha .$$

We leave it to the reader to verify that if

$$\sum_{\alpha\in\mathcal{I}} x_\alpha < \infty \ ,$$

then at most countably many x_α are nonzero.

THEOREM 3.15 (Bessel's inequality). *Let H be a Hilbert space and $\{u_\alpha\}_{\alpha\in\mathcal{I}} \subset H$ an* ON *set. For $x \in H$,*

$$\sum_{\alpha\in\mathcal{I}} |(x,u_\alpha)|^2 \le \|x\|^2 \ .$$

PROOF. By the previous theorem, for any finite $\mathcal{J} \subset \mathcal{I}$,

$$\sum_{\alpha\in\mathcal{J}} |(x,u_\alpha)|^2 \le \|x\|^2 \ ,$$

so the same is true of the supremum. □

COROLLARY 3.16. *At most countably many of the (x,u_α) are nonzero.*

In a sense to be made precise below in the Riesz-Fischer Theorem, $x \in H$ can be associated to its coefficients $(x,u_\alpha)\ \forall\ \alpha$. We define a space of coefficients below.

DEFINITION. Let $\mathcal{I}$ be a set. We denote by $\ell_2(\mathcal{I})$ the set

$$\ell_2(\mathcal{I}) = \{f : \mathcal{I} \to \mathbb{F} : \sum_{\alpha\in\mathcal{I}} |f(\alpha)|^2 < \infty\} \ .$$

If $\mathcal{I} = \mathbb{N}$, we have the usual space ℓ_2, which is a Hilbert space. In general, we have an inner-product on $\ell_2(\mathcal{I})$ given by

$$(f,g) = \sum_{\alpha\in\mathcal{I}} f(\alpha)\overline{g(\alpha)} \ ,$$

as the reader can verify. Moreover, $\ell_2(\mathcal{I})$ is complete.

THEOREM 3.17 (Riesz-Fischer Theorem). *Let H be a Hilbert space and $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ any* ON *set in H. Define the mapping $F : H \to \ell_2(\mathcal{I})$ by $F(x) = f_x$ where*

$$f_x(\alpha) = x_\alpha \equiv (x,u_\alpha)$$

for $\alpha \in \mathcal{I}$. Then F is a surjective bounded linear map.

PROOF. Denoting the map f_x by $\{x_\alpha\}_{\alpha\in\mathcal{I}}$, the mapping F is linear since

$$\begin{aligned} F(x+y) &= \{(x+y)_\alpha\}_{\alpha\in\mathcal{I}} = \{(x+y,u_\alpha)\}_{\alpha\in\mathcal{I}} \\ &= \{(x,u_\alpha)+(y,u_\alpha)\}_{\alpha\in\mathcal{I}} \\ &= \{(x,u_\alpha)\}_{\alpha\in\mathcal{I}} + \{(y,u_\alpha)\}_{\alpha\in\mathcal{I}} \\ &= F(x)+F(y) \end{aligned}$$

and similarly for scalar multiplication. F is a bounded map because of Bessell's inequality

$$\|F(x)\|^2_{\ell_2(\mathcal{I})} = \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 \le \|x\|^2_H \ .$$

Thus, not only is F bounded, but

$$\|F\|_{B(H,\ell_2(\mathcal{I}))} \le 1 \ .$$

The interesting point is that F is surjective. Let $f \in \ell_2(\mathcal{I})$ and let $n \in \mathbb{N}$. If

$$\mathcal{I}_n = \left\{\alpha \in \mathcal{I} : |f(\alpha)| > \frac{1}{n}\right\} ,$$

then if $|\mathcal{I}_n|$ denotes the number of α in $\mathcal{I}_n$,

$$|\mathcal{I}_n| \leq n^2 \|f\|^2_{\ell_2(\mathcal{I})} .$$

Let $\mathcal{J} = \bigcup_{n=1}^\infty \mathcal{I}_n$. Then J is countable and if $\beta \notin \mathcal{J}$, then $f(\beta) = 0$. In H, define x_n by

$$x_n = \sum_{\alpha \in \mathcal{I}_n} f(\alpha) u_\alpha .$$

Since $\mathcal{I}_n$ is a finite set, x_n is a well-defined element of H. We expect that $\{x_n\}_{n=1}^\infty$ is Cauchy in H. To see this, let $n > m \geq 1$ and compute

$$\|x_n - x_m\|^2 = \Big\| \sum_{\alpha \in \mathcal{I}_n \smallsetminus \mathcal{I}_m} f(\alpha) u_\alpha \Big\|^2 = \sum_{\alpha \in \mathcal{I}_n \smallsetminus \mathcal{I}_m} |f(\alpha)|^2 \leq \sum_{\alpha \in \mathcal{I} \smallsetminus \mathcal{I}_m} |f(\alpha)|^2$$

and the latter is the tail of an absolutely convergent series, and so is as small as we like provided we take m large enough. Since H is a Hilbert space, there is an $x \in H$ such that $x_n \xrightarrow{H} x$. As F is continuous, $F(x_n) \to F(x)$. We show that $F(x) = f$. By continuity of the inner-product, for $\alpha \in \mathcal{I}$

$$\begin{aligned} F(x)(\alpha) = (x, u_\alpha) &= \lim_{n \to \infty} (x_n, u_\alpha) \\ &= \lim_{n \to \infty} \sum_{\beta \in \mathcal{I}_n} f(\beta)(u_\beta, u_\alpha) = f(\alpha) . \qquad \square \end{aligned}$$

THEOREM 3.18. *Let H be a Hilbert space. The following are equivalent conditions on an* ON *set $\{u_\alpha\}_{\alpha \in \mathcal{I}} \subset H$.*

(i) $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ *is a maximal* ON *set (also called an* ON basis *for H).*
(ii) $\mathrm{Span}\{u_\alpha : \alpha \in \mathcal{I}\}$ *is dense in H.*
(iii) $\|x\|_H^2 = \sum_{\alpha \in \mathcal{I}} |(x, u_\alpha)|^2$ *for all $x \in H$.*
(iv) $(x, y) = \sum_{\alpha \in \mathcal{I}} (x, u_\alpha)\overline{(y, u_\alpha)}$ *for all $x, y \in H$.*

PROOF. (i) $\implies$ (ii). Let $M = \overline{\mathrm{span}\{u_\alpha\}}$. Then M is a closed linear subspace of H. If M is not all of H, $M^\perp \neq \{0\}$ since $H = M + M^\perp$. Let $x \in M^\perp$, $x \neq 0$, $\|x\| = 1$. Then the set $\{u_\alpha : \alpha \in \mathcal{I}\} \cup \{x\}$ is an ON set, so $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ is not maximal, a contradiction.

(ii) $\implies$ (iii). We are assuming $M = H$ in the notation of the last paragraph. Let $x \in H$. Because of Bessell's inequality,

$$\|x\|^2 \geq \sum_{\alpha \in \mathcal{I}} |x_\alpha|^2 ,$$

where $x_\alpha = (x, u_\alpha)$ for $\alpha \in \mathcal{I}$. Let $\varepsilon > 0$ be given. Since $\mathrm{span}\{u_\alpha : \alpha \in \mathcal{I}\}$ is dense, there is a finite set $\alpha_1, \dots, \alpha_N$ and constants $c_1, \dots, c_N$ such that

$$\Big\| x - \sum_{i=1}^N c_i u_{\alpha_i} \Big\| \leq \varepsilon .$$

By the Best Approximation analysis, on the other hand,

$$\Big\| x - \sum_{i=1}^N x_{\alpha_i} u_{\alpha_i} \Big\| \leq \Big\| x - \sum_{i=1}^N c_i u_{\alpha_i} \Big\| .$$

It follows from orthonormality of the $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ that

$$\varepsilon^2 \geq \Big\| x - \sum_{i=1}^N x_{\alpha_i} u_{\alpha_i} \Big\|^2 = \|x\|^2 - \sum_{i=1}^N |x_{\alpha_i}|^2 \geq \|x\|^2 - \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 \ .$$

In consequence,

$$\|x\|^2 \leq \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 + \varepsilon \ ,$$

and $\varepsilon > 0$ was arbitrary. Thus equality holds everywhere in Bessell's inequality.

(iii) $\implies$ (iv). This follows because in a Hilbert space, the norm determines the inner-product as we now show. Let $x, y \in H$. Because of (iii), we have

$$\begin{aligned}\|x\|^2 + \|y\|^2 + (x,y) + (y,x) &= \|x+y\|^2 \\ &= \sum_{\alpha\in\mathcal{I}} |x_\alpha + y_\alpha|^2 = \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 + \sum_{\alpha\in\mathcal{I}} |y_\alpha|^2 + \sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha + \sum_{\alpha\in\mathcal{I}} \bar{x}_\alpha y_\alpha \ ;\end{aligned}$$

whereas

$$\begin{aligned}\|x\|^2 + \|y\|^2 + i(y,x) - i(x,y) &= \|x+iy\|^2 \\ &= \sum_{\alpha\in\mathcal{I}} |x_\alpha + iy_\alpha|^2 = \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 + \sum_{\alpha\in\mathcal{I}} |y_\alpha|^2 + i\sum_{\alpha\in\mathcal{I}} y_\alpha \bar{x}_\alpha - i\sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha \ .\end{aligned}$$

Since

$$\|x\|^2 = \sum_{\alpha\in\mathcal{I}} |x_\alpha|^2 \quad \text{and} \quad \|y\|^2 = \sum_{\alpha\in\mathcal{I}} |y_\alpha|^2 \ ,$$

it is ascertained that

$$(x,y) + \overline{(x,y)} = \sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha + \overline{\sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha}$$

and

$$(x,y) - \overline{(x,y)} = \sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha - \overline{\sum_{\alpha\in\mathcal{I}} x_\alpha \bar{y}_\alpha} \ ,$$

and the desired result follows.

(iv) $\implies$ (i). If $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is not a maximal ON set, let $u \in H$, $u \perp u_\alpha$ for all $\alpha \in \mathcal{I}$, and $\|u\| = 1$. Then, because of (iv),

$$1 = \|u\|^2 = \sum_{\alpha\in\mathcal{I}} |(u, u_\alpha)|^2 = 0 \ ,$$

a contradiction. □

Corollary 3.19. *If $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is maximal* ON *and $x \in H$ is infinite dimensional, then there are $\alpha_i \in \mathcal{I}$ for $i = 1, 2, \ldots$ such that*

$$x = \sum_{i=1}^\infty (x, u_{\alpha_i}) u_{\alpha_i} \ .$$

Proof. Exercise. □

That is, indeed, a maximal ON set is a type of basis for the Hilbert space.

Corollary 3.20. *If $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is a maximal* ON *set, then the Riesz-Fischer map $F : H \to \ell_2(\mathcal{I})$ is a Hilbert space isomorphism.*

THEOREM 3.21. *Let H be a Hilbert space and $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ any* ON *set in H. Then $\{u_\alpha\}_{\alpha\in\mathcal{I}} \subset \{u_\beta\}_{\beta\in\mathcal{J}}$ where the latter is* ON *and maximal.*

PROOF. The general result follows from transfinite induction. We prove the result assuming that H is also separable.

Let $\{\tilde{x}_j\}_{j=1}^\infty$ be dense in H and

$$M = \overline{\operatorname{span}\{u_\alpha\}_{\alpha\in\mathcal{I}}} \ .$$

Define

$$\hat{x}_j = \tilde{x}_j - P_M\tilde{x}_j \in M^\perp \ ,$$

where P_M is orthogonal projection onto M. Then the span of

$$\{u_\alpha\}_{\alpha\in\mathcal{I}} \cup \{\hat{x}_j\}_{j=1}^\infty$$

is dense in H. Define successively for $j = 1, 2, \ldots$ (with $x_1 = \hat{x}_1$)

$$N_j = \overline{\operatorname{span}\{x_1, \ldots, x_j\}} \ ,$$

$$x_{j+1} = \hat{x}_{j+1} - P_{N_j}\hat{x}_{j+1} \in N_j^\perp \ .$$

Then the span of

$$\{u_\alpha\}_{\alpha\in\mathcal{I}} \cup \{x_j\}_{j=1}^\infty$$

is dense in H and any two elements are orthogonal. Remove any zero vectors and normalize to complete the proof by the equivalence of (ii) and (iii) in Theorem 3.18. $\square$

COROLLARY 3.22. *Every Hilbert space H is isomorphic to $\ell_2(\mathcal{I})$ for some $\mathcal{I}$. Moreover, H is infinite dimensional and separable if and only if H is isomorphic to $\ell_2(\mathbb{N})$.*

We illustrate orthogonality in a Hilbert space by considering Fourier series. If $f : \mathbb{R} \to \mathbb{C}$ is periodic of period T, then $g : \mathbb{R} \to \mathbb{C}$ defined by $g(x) = f(\lambda x)$ for some $\lambda \neq 0$ is periodic of period T/λ. So when considering periodic functions, it is enough to restrict to the case $T = 2\pi$.

Let

$$L_{2,\text{per}}(-\pi,\pi) = \{f : \mathbb{R} \to \mathbb{C} : f \in L_2([-\pi,\pi)) \text{ and } f(x + 2n\pi) = f(x) \\ \text{for a.e. } x \in [-\pi,\pi) \text{ and integer } n\} \ .$$

With the inner-product

$$(f,g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,\overline{g(x)}\,dx \ ,$$

$L_{2,\text{per}}(-\pi,\pi)$ is a Hilbert space (it is left to the reader to verify these assertions). The set

$$\{e^{inx}\}_{n=-\infty}^\infty \subset L_{2,\text{per}}(-\pi,\pi)$$

is ON, as can be readily verified.

THEOREM 3.23. *The set* $\operatorname{span}\{e^{inx}\}_{n=-\infty}^\infty$ *is dense in $L_{2,\text{per}}(-\pi,\pi)$.*

PROOF. We first remark that $C_{\text{per}}([-\pi,\pi])$, the continuous functions defined on $[-\infty,\infty]$ that are periodic, are dense in $L_{2,\text{per}}(-\pi,\pi)$. This follows from the fact that simple functions are dense in L_2 (their limits are used to define the integral). By "rounding out the corners", in a manner to be made precise when we study distributions and convolutions, we can show the density of $C_{\text{per}}([-\pi,\pi])$ in $L_{2,\text{per}}(-\pi,\pi)$. Thus it is enough to show that a continuous and periodic function f of period 2π is the limit of functions in $\operatorname{span}\{e^{inx}\}_{n=-\infty}^\infty$.

For any integer $m \geq 0$, on $[-\pi, \pi]$ let

$$k_m(x) = c_m \left(\frac{1+\cos x}{2}\right)^m \geq 0$$

where c_m is defined so that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} k_m(x)\,dx = 1 \ . \tag{3.2}$$

As $m \to \infty$, $k_m(x)$ is concentrated about $x = 0$ but maintains total integral 2π (i.e., $k_m/2\pi \to \delta_0$, the Dirac distribution to be defined later). Now

$$k_m(x) = c_m \left[\frac{2 + e^{ix} + e^{-ix}}{4}\right]^m \in \text{span}\ \{e^{inx}\}_{n=-m}^m \ ,$$

and so, for some $\lambda_n \in \mathbb{C}$,

$$\begin{aligned}
f_m(x) &\equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} k_m(x-y) f(y)\,dy \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{n=-m}^{m} \lambda_n e^{in(x-y)} f(y)\,dy \\
&= \sum_{n=-m}^{m} \left(\frac{\lambda_n}{2\pi}\int_{-\pi}^{\pi} e^{-iny} f(y)\,dy\right) e^{inx} \in \text{span}\ \{e^{inx}\}_{n=-m}^m \ .
\end{aligned}$$

We claim that in fact $f_n \to f$ uniformly in the L_∞ norm, so also in L_2, and the proof will be complete. By periodicity,

$$f_m(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y) k_m(y)\,dy \ ,$$

and, by (3.2),

$$f(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x) k_m(y)\,dy \ .$$

Thus, for any $\delta > 0$,

$$\begin{aligned}
|f_m(x) - f(x)| &= \frac{1}{2\pi}\left|\int_{-\pi}^{\pi} \Big(f(x-y) - f(x)\Big) k_m(y)\,dy\right| \\
&\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x-y) - f(x)| k_m(y)\,dy \\
&= \frac{1}{2\pi}\int_{\delta<|y|\leq\pi} |f(x-y) - f(x)| k_m(y)\,dy \\
&\qquad + \frac{1}{2\pi}\int_{|y|\leq\delta} |f(x-y) - f(x)| k_m(y)\,dy \ .
\end{aligned}$$

Given $\varepsilon > 0$, since f is continuous on $[-\pi, \pi]$, it is uniformly continuous. Thus there is $\delta > 0$ such that $|f(x-y) - f(x)| < \varepsilon/2$ for all $|y| \leq \delta$, and the last term on the right side above is bounded by $\varepsilon/2$. For the next to last term, we note that from (3.2),

$$\begin{aligned} 1 &= \frac{c_m}{\pi}\int_0^\pi \left(\frac{1+\cos x}{2}\right)^m \,dx \\ &\geq \frac{c_m}{\pi}\int_0^\pi \left(\frac{1+\cos x}{2}\right)^m \sin x\,dx \\ &= \frac{c_m}{(m+1)\pi}\ , \end{aligned}$$

which implies that

$$c_m \leq (m+1)\pi\ .$$

Now f is continuous on $[-\pi,\pi]$, so there is $M \geq 0$ such that $|f(x)| \leq M$. Thus for $|y| > \delta$,

$$k_m(y) \leq (1+m)\pi\left(\frac{1+\cos\delta}{2}\right)^m < \frac{\varepsilon}{4M}$$

for m large enough. Combining, we have that

$$|f_m(x) - f(x)| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi} 2M\frac{\varepsilon}{4M}\,dy + \frac{\varepsilon}{2} = \varepsilon\ .$$

We conclude that $f_n \xrightarrow{L_\infty} f$ uniformly. □

3.5. Weak Convergence in a Hilbert Space

Because of the Riesz Representation Theorem, a sequence $\{x_n\}_{n=1}^\infty$ from a Hilbert space H converges weakly to x is and only if

$$(x_n, y) \longrightarrow (x, y) \tag{3.3}$$

for all $y \in H$. In fact, if $\{e_\alpha\}_{\alpha\in\mathcal{I}}$ is an ON base for H, then $x_n \xrightarrow{w} x$ if and only if the Fourier coefficients

$$(x_n, e_\alpha) \xrightarrow{n\to\infty} (x, e_\alpha) \tag{3.4}$$

for all $\alpha \in \mathcal{I}$.

Clearly (3.3) implies (3.4). On the other hand suppose (3.4) is valid and let $y \in H$. Since $\{e_\alpha\}_{\alpha\in\mathcal{I}}$ is an ON base, we know from the Riesz-Fischer Theorem that $\operatorname{span}\{e_\alpha\}_{\alpha\in\mathcal{I}}$ is dense in H. Let $\varepsilon > 0$ be given and let $\{c_\alpha\}_{\alpha\in\mathcal{I}}$ be a collection of constants such that $c_\alpha = 0$ for all but a finite number of α and so that $z = \sum_{\alpha\in\mathcal{I}} c_\alpha e_\alpha \in \operatorname{span}\{e_\alpha\}_{\alpha\in\mathcal{I}}$ satisfies

$$\|y - z\| < \varepsilon\ .$$

Because of (3.4),

$$(x_n, z) \xrightarrow{n\to\infty} (x, z)$$

since z is a finite linear combination of the e_α's. But then,

$$\begin{aligned} \limsup_{n\to\infty} |(x_n - x, y)| &\leq \limsup_{n\to\infty} |(x_n - x, y - z)| + \limsup_{n\to\infty} |(x_n - x, z)| \\ &= \limsup_{n\to\infty} |(x_n - x, y - z)| \\ &\leq \left(\sup_{n\geq 1}\|x_n\| + \|x\|\right)\|y - z\| \\ &\leq M\varepsilon\ . \end{aligned}$$

It follows that

$$\lim_{n\to\infty} |(x_n - x, y)| \le M\varepsilon \ ;$$

and as $\varepsilon > 0$ was arbitrary and, provided M is finite and does not depend up ε, this means that

$$\lim_{n\to\infty} (x_n, y) = (x, y) \ ,$$

as required.

On the other hand $\{x_n\}_{n=1}^{\infty}$ weakly convergent implies it to be weakly bounded, and hence by the Uniform Boundedness Principle, $\{\alpha_n\}_{n=1}^{\infty}$ is bounded in norm.

Example. Consider $L_2(-\pi, \pi)$ and consider the ON set

$$\{e^{inx}\}_{n=-\infty}^{\infty} \ .$$

This sequence converges weakly to zero, for obviously if m is fixed,

$$(e^{inx}, e^{imx}) = 0$$

for $n > m$. However, as $\|e^{inx} - e^{imx}\| = \sqrt{2}$ for $n \ne m$, the sequence is *not* Cauchy in norm, and so has no strong limit.

3.6. Basic spectral theory in Banach spaces

We turn now to a discussion of spectral theory, which is concerned with questions of invertibility of an operator. Initially our theory will be developed for operators in any Banach space; later we will restrict to Hilbert spaces. So let X be a complex Banach space (so $\mathbb{F} = \mathbb{C}$) and $T \in B(X, X)$ a bounded linear operator. The range or image of T is $R(T) \subset X$, and the null space or kernel is

$$N(T) = \{x \in X : Tx = 0\} \subset X \ .$$

For $\lambda \in \mathbb{C}$, we consider

$$T_\lambda = T - \lambda I \ ,$$

where I is the identity operator on X. There are two possibilities. Either T_λ is one-to-one ($N(T) = \{0\}$) and onto ($R(T) = X$), i.e., T_λ invertible, or it is not.

Definition. If T_λ is invertible, then λ is said to be in the *resolvent set* of T, denoted $\rho(T) \subset \mathbb{C}$. That is,

$$\lambda \in \rho(T) = \{\mu \in \mathbb{C} : T_\mu = T - \mu I \text{ is one-to-one and onto}\} \ .$$

Also, T_λ^{-1} is then called the *resolvent operator*.

Proposition 3.24. *If $\lambda \in \rho(T)$, then $T_\lambda^{-1} \in B(X, X)$.*

Proof. This follows from the open mapping theorem. □

If $\lambda \in \mathbb{C}$ is not in $\rho(T)$, then T_λ is not invertible. In infinite dimensions, there are several possibilities for why λ fails to lie in $\rho(T)$.

Definition. If $\lambda \notin \rho(T)$, then we say that λ lies in the *spectrum* of T. We denote the spectrum of T by

$$\sigma(T) = \mathbb{C} \setminus \rho(T) = \{\mu \in \mathbb{C} : T_\mu \text{ is not both one-to-one and onto}\} \ ,$$

and subdivide it into the *point spectrum* of T,

$$\sigma_p(T) = \{\mu \in \mathbb{C} : T_\mu \text{ is not one-to-one}\} ,$$

the *continuous spectrum* of T,

$$\sigma_c(T) = \{\mu \in \mathbb{C} : T_\mu \text{ is one-to-one and } R(T_\mu) \text{ is dense in } X, \text{ but } T_\mu^{-1} \text{ is not bounded}\} ,$$

and the *residual spectrum* of T,

$$\sigma_r(T) = \{\mu \in \mathbb{C} : T_\mu \text{ is one-to-one and } R(T_\mu) \text{ is not dense in } X\} .$$

PROPOSITION 3.25. *The point, continuous, and residual spectra are disjoint and their union is $\sigma(T)$.*

PROOF. We need only show that if T_μ is one-to-one, has dense range, and T_μ^{-1} is bounded, then $\mu \in \rho(T)$, i.e., T_μ is onto. If so, the proposition is obvious.

Let $S = T_\mu^{-1} : R(T_\mu) \to X$, a bounded linear operator. We note that by density of $R(T_\mu)$ and completeness of X, S extends to $\tilde{S} \in B(X, X)$, defined by

$$\tilde{S}y = \lim_{n\to\infty} Sy_n$$

for any $y \in X$ and $y_n \to y$ with $\{y_n\}_{n=1}^\infty \subset R(T_\mu)$ (the reader should verify that indeed $\tilde{S}$ so defined is in $B(X, X)$).

Now for any such y and $\{y_n\}_{n=1}^\infty$,

$$\tilde{S}y_n = Sy_n = T_\mu^{-1}y_n \equiv x_n \in X .$$

but then $x_n = \tilde{S}y_n \to \tilde{S}y \equiv x \in X$, so, since T_μ is continuous,

$$y_n = T_\mu x_n \to T_\mu x .$$

Thus $T_\mu x = y$, and T_μ is onto. □

If $\lambda \in \sigma_p(T)$, then

$$N(T_\lambda) \neq \{0\} ,$$

so there are $x \in X$, $x \neq 0$, such that $T_\lambda x = 0$; that is,

$$Tx = \lambda_x .$$

DEFINITION. The complex numbers in $\sigma_p(T)$ are called *eigenvalues*, and any $x \in X$ such that $x \neq 0$ and

$$Tx = \lambda x$$

is called an *eigenfunction* or *eigenvector* of T corresponding to $\lambda \in \sigma_p(T)$.

LEMMA 3.26. *Let X be a Banach space and $V \in B(X, X)$ with $\|V\| < 1$. Then $I - V \in B(X, X)$ is one-to-one and onto, hence by the open mapping theorem has a bounded inverse.*

PROOF. Let $N > 0$ be an integer and let

$$S_N = I + V + V^2 + \cdots + V^N = \sum_{n=0}^{N} V^n .$$

Then $S_N \in B(X,X)$ for all N. The sequence $\{S_N\}_{N=1}^\infty$ is Cauchy in $B(X,X)$, for if $M > N$, then

$$\|S_M - S_N\|_{B(X,X)} = \Big\| \sum_{n=N+1}^{M} V^n \Big\|_{D(X,X)} \le \sum_{n=N+1}^{M} \|V\|_{B(X,X)}^n$$

and this tends to zero as $N \to \infty$ since $\mu = \|V\|_{B(X,X)} < 1$ implies $\sum_{k=0}^\infty \mu^k < +\infty$. Since $B(X,X)$ is a Banach space, it follows that there is an $S \in B(X,X)$ such that $S_N \to S$.

We now show that $(I - V)S = S(I - V) = I$. First, for each $N = 1, 2, \dots,$ obviously $VS_N = S_N V$ and hence V commutes with S. Second, notice that

$$VS_N = S_{N+1} - I = V^{N+1} + S_N - I \ .$$

Rearranging gives

$$(I - V)S_N = I - V^{N+1} \ . \tag{3.5}$$

On the other hand $V^{N+1} \to 0$ in $B(X,X)$ since

$$\|V^{N+1}\|_{B(X,X)} \le \|V\|_{B(X,X)}^{N+1} \to 0$$

as $N \to \infty$. Since $S_N \to S$ in $B(X,X)$, it follows readily that $TS_N \to TS$ for any $T \in B(X,X)$. Thus we may take the limit as $N \to +\infty$ in (3.5) to obtain

$$(I - V)S = I = S(I - V) \ ,$$

the latter since V and S commute. These two relations imply $I - V$ to be onto and one-to-one, respectively. $\square$

Corollary 3.27. *If V is as above, then*

$$(I - V)^{-1} = \sum_{n=0}^\infty V^n \ .$$

The latter expression is called the Neumann series *for V.*

Corollary 3.28. *Let X be a Banach space. Then the set of invertible operators in $B(X,X)$ is open.*

Proof. Let $A \in B(X,X)$ be such that $A^{-1} \in B(X,X)$. Let $\varepsilon > 0$ be such that $\varepsilon \le 1/\|A^{-1}\|_{B(X,X)}$. Choose any $B \in B(X,X)$ with $\|B\| < \varepsilon$. Then $A + B$ is invertible. To see this, write

$$A + B = A(I + A^{-1}B)$$

and note that

$$\|A^{-1}B\|_{B(X,X)} \le \|A^{-1}\|_{B(X,X)} \|B\|_{B(X,X)} < \varepsilon \|A^{-1}\|_{B(X,X)} < 1 \ .$$

Hence $I + A^{-1}B$ is boundedly invertible, and thus so is $A(I + A^{-1}B)$ since it is a composition of two invertible operators. $\square$

Corollary 3.29. *Let $T \in B(X,X)$. Then $\rho(T)$ is an open subset of $\mathbb{C}$.*

Proof. If $\lambda \in \rho(T)$, then $T - \lambda I$ is invertible. Hence $T - \lambda I + B$ is invertible of $\|B\|_{B(X,X)}$ is small enough. In particular,

$$T - \lambda I - \mu I$$

is invertible if $|\mu|$ is small enough. Thus $\lambda \in \rho(T)$ implies $\lambda + \mu \in \rho(T)$ if $|\mu|$ is small enough, and so $\rho(T)$ is open. $\square$

Corollary 3.30. *Suppose X is a Banach space and $T \in B(X,X)$. If $R = \|T\|_{B(X,X)}$, then*

$$\sigma(T) \subseteq \overline{B_R(0)} \ .$$

Proof. We show that if $|\lambda| > R$, then $\lambda \in \rho(T)$. But this is straightforward since

$$T - \lambda I = -\lambda\left(I - \frac{1}{\lambda}T\right)$$

and $\|\frac{1}{\lambda}T\| = \frac{1}{|\lambda|}\|T\| < 1$. $\square$

Corollary 3.31. *Let X be a Banach space and $T \in B(X,X)$. Then $\sigma(T)$ is compact.*

Proof. It is closed and bounded. $\square$

We should caution the reader that we have not shown that $\sigma(T) \neq \emptyset$; such is possible. To continue, we will restrict to certain classes of operators where we can say more.

Remark. Although we have required $T \in B(X,X)$, much of the theory can be developed for unbounded linear operators that are *densely defined*, that is, for a linear operator T with domain of definition $D(T) \subset X$ dense in X, and with range $R(T) \subset X$. However, strange things can happen. For example, let $X = L_2(-1,1)$ and $T = d/dx$. Then $D(T) = C^1(-1,1)$, say, which is dense in X, but T is unbounded (consider $f(x) = \sin nx \in X$). Let $\lambda \in \mathbb{C}$ and note that

$$(T - \lambda I)e^{\lambda x} = 0 \ .$$

Hence every $\lambda \in \mathbb{C}$ is an eigenvalue, $\sigma(T) = \sigma_p(T) = \mathbb{C}$, and $\rho(T) = \emptyset$.

3.7. Bounded self-adjoint linear operators

We return now to an operator $T \in B(H,H)$ defined on a Hilbert space H. Because of the Riesz representation theorem, the adjoint operator $T^* : H^* \to H^*$ is also defined on $H \cong H^*$. That is, we consider that $T^* \in B(H,H)$. In this case, we call T^* the *Hilbert-adjoint* operator for T. Let us consider its action. If $L_y \in H^*$ for some $y \in H$ and $x \in H$, then, by definition,

$$(T^* L_y)(x) = L_y(Tx) = (Tx, y) \ .$$

Now $T^* L_y = L_z$ for some $z \in H$. Call $z = T^* y$, and then $T^* L_y = L_{T^* y}$, so

$$(x, T^* y) = (Tx, y) \ \forall \ x, y \in H \ .$$

Proposition 3.32. *Let H be a Hilbert space and $T \in B(H,H)$. Then $T = T^{**}$ and $(T^* x, y) = (x, Ty) \ \forall \ x, y \in H$.*

Proof. Exercise. $\square$

We consider maps T for which $T = T^*$.

Definition. If H is a Hilbert space, $T \in B(H,H)$, and $T = T^*$ (interpreted as above), then T is said to be *self-adjoint* or *Hermitian*.

Proposition 3.33. *Let H be a Hilbert space and $T \in B(H,H)$.*

(a) *If T is self-adjoint, then*

$$(Tx, x) \in \mathbb{R} \quad \forall\, x \in H \ .$$

(b) *If H is a complex Hilbert space, then T is self-adjoint if and only if (Tx, x) is real for all $x \in H$.*

Proof. (a) We compute

$$(Tx, x) = \overline{(x, Tx)} = \overline{(x, T^*x)} = \overline{(Tx, x)} \in \mathbb{R} \ .$$

(b) By (a), we need only show the converse. This will follow if we can show that

$$(Tx, y) = (T^*x, y) \quad \forall\, x, y \in H \ .$$

Let $\alpha \in \mathbb{C}$ and compute

$$\begin{aligned} \mathbb{R} \ni & \Big(T(x + \alpha y), x + \alpha y\Big) \\ &= (Tx, x) + |\alpha|^2 (Ty, y) + \alpha(Ty, x) + \bar\alpha(Tx, y) \ . \end{aligned}$$

The first two terms on the right are real, so also the sum of the latter two. Thus

$$\mathbb{R} \ni \bar\alpha(Tx, y) + \overline{\bar\alpha(T^*x, y)} \ .$$

If $\alpha = 1$, we conclude that the complex parts of (Tx, y) and (T^*x, y) agree; if $\alpha = i$, the real parts agree. ☐

We isolate an important result that is useful in other contexts.

Lemma 3.34. *Suppose X and Y are Banach spaces and $T \in B(X, Y)$. Suppose that T is* bounded below*, i.e., there is some $\gamma > 0$ such that*

$$\|Tx\|_Y \ge \gamma \|x\|_X \quad \forall\, x \in X \ .$$

Then T is one-to-one and $R(T)$ is closed in Y.

Proof. That T is one-to-one is clear by linearity. Suppose for $n = 1, 2, \dots$, $y_n = Tx_n$ is a sequence in $R(T)$ and that $y_n \to y \in Y$. Then $\{y_n\}_{n=1}^\infty$ is Cauchy, so also is $\{x_n\}_{n=1}^\infty$. Since X is complete, there is $x \in X$ such that $x_n \to x$. Since T is continuous, $y_n = Tx_n \to Tx = y \in R(T)$; that is, $R(T)$ is closed. ☐

Theorem 3.35. *Let H be a Hilbert space and $T \in B(H,H)$ self-adjoint. Then $\sigma_p(T) \subset \mathbb{R}$. Moreover, $\lambda \in \rho(T)$ if and only if there is some $\gamma > 0$ such that*

$$\|T_\lambda x\| \ge \gamma \|x\| \quad \forall\, x \in H \ .$$

Proof. If $\lambda \in \sigma_p(T)$ and $Tx = \lambda x$ for $x \neq 0$, then $\lambda(x, x) = (Tx, x) = (x, Tx) = (x, \lambda x) = \bar\lambda(x, x)$; thus $\lambda = \bar\lambda$ is real.

If $\lambda \in \rho(T)$, then the final conclusion follows from the boundedness of T_λ^{-1},

$$\|x\| = \|T_\lambda^{-1} T_\lambda x\| \le \|T_\lambda^{-1}\| \, \|T_\lambda x\| \ ,$$

and the fact that $T_\lambda^{-1} \not\equiv 0$.

Finally, suppose T_λ is bounded below. By the lemma, T_λ is one-to-one and $R(T_\lambda)$ is closed. If $R(T_\lambda) \neq H$, then there is some $x_0 \in R(T_\lambda)^\perp$, and, $\forall\ x \in H$,

$$\begin{aligned} 0 = (T_\lambda x, x_0) &= (Tx, x_0) - \lambda(x, x_0) \\ &= (x, Tx_0) - \lambda(x, x_0) \\ &= (x, T_{\bar\lambda} x_0)\ . \end{aligned}$$

Thus $T_{\bar\lambda} x_0 = 0$, or $Tx_0 = \bar\lambda x_0$ and $\bar\lambda \in \sigma_p(T)$. But then $\lambda = \bar\lambda \in \sigma_p(T)$, and T_λ is not one-to-one, a contradiction. Thus $R(T_\lambda) = H$ and $\lambda \in \rho(T)$. □

Corollary 3.36. *The spectrum $\sigma(T)$ of a bounded self-adjoint linear operator T on a Hilbert space H is real.*

Proof. Suppose $\lambda = \alpha + i\beta \in \sigma(T)$, where $\alpha, \beta \in \mathbb{R}$. For any $x \neq 0$ in H,

$$(T_\lambda x, x) = (Tx, x) - \lambda(x, x)$$

and

$$\overline{(T_\lambda x, x)} = (Tx, x) - \bar\lambda(x, x)\ ,$$

since (Tx, x) is real. Thus

$$(T_\lambda x, x) - \overline{(T_\lambda x, x)} = 2i\beta(x, x)\ ,$$

or

$$\beta\|x\|^2 = \frac{1}{2i}\Big[(T_\lambda x, x) - \overline{(T_\lambda x, x)}\Big] \leq \|T_\lambda x\|\,\|x\|\ .$$

As $x \neq 0$, we see that if $\beta \neq 0$, T_λ is bounded below, and conclude $\lambda \in \rho(T)$, a contradiction. □

Corollary 3.37. *The residual spectrum $\sigma_r(T)$ of a bounded self-adjoint operator T on a Hilbert space H is empty.*

Proof. Suppose not. Let $\lambda \in \sigma_r$. Then T_λ is invertible on its range

$$T_\lambda^{-1} : R(T_\lambda) \to H\ ,$$

but

$$\overline{R(T_\lambda)} \neq H\ .$$

Let

$$y \in \overline{R(T_\lambda)}^\perp \setminus \{0\}\ .$$

Then, $\forall\ x \in H$,

$$0 = (T_\lambda x, y) = (x, T_\lambda y)\ .$$

Let $x = T_\lambda y$ to conclude that $T_\lambda y = 0$, i.e., $\lambda \in \sigma_p(T)$. Since $\sigma_r(T) \cap \sigma_p(T) = \emptyset$, we have our contradiction. □

Thus the spectrum of T is real and consists only of eigenvalues ($\sigma_p(T)$) and the continuous spectrum. In fact we can bound $\sigma(T)$ on the real line.

PROPOSITION 3.38. *Let H be a Hilbert space and $T \in B(H,H)$ be a self-adjoint operator. Then*

$$\sigma(T) \subset [r, R]$$

where

$$r = \inf_{\|x\|=1} (Tx, x) \quad \text{and} \quad R = \sup_{\|x\|=1} (Tx, x) \ .$$

PROOF. Let $c > 0$ and let $\lambda = R + c > R$. Let $x \neq 0$ and compute

$$(Tx, x) = \|x\|^2 \left(T\Big(\frac{x}{\|x\|}\Big), \frac{x}{\|x\|} \right) \leq \|x\|^2 R \ .$$

On the other hand,

$$-(Tx - \lambda x, x) = -(T_\lambda x, x) \leq \|T_\lambda x\| \, \|x\| \ ,$$

and

$$-(Tx - \lambda x, x) = -(Tx, x) + \lambda \|x\|^2 \geq -\|x\|^2 R + \lambda \|x\|^2 = c\|x\|^2 \ .$$

It is concluded that

$$\|T_\lambda x\| \geq c\|x\| \ ;$$

hence, $\lambda \in \rho(T)$.

A similar argument applies in case $\lambda = r - c$ where $c > 0$. Write for $x \neq 0$

$$(Tx, x) = \|x\|^2 \left(T\Big(\frac{x}{\|x\|}\Big), \frac{x}{\|x\|} \right) \geq \|x\|^2 r \ .$$

On the other hand,

$$(Tx - \lambda x, x) = (T_\lambda x, x) \leq \|T_\lambda x\| \, \|x\| \ ,$$

and

$$(Tx - \lambda x, x) = (Tx, x) - \lambda \|x\|^2 \geq (r - \lambda)\|x\|^2 = c\|x\|^2 \ ,$$

so $\lambda \in \rho(T)$. $\square$

We call

$$q(x) = \frac{(Tx, x)}{(x, x)} \quad \forall \, x \neq 0$$

the *Rayleigh quotient* of T at x. The result above is that

$$\sigma(T) \subset \left[\inf_{x \neq 0} q(x) \ , \ \sup_{x \neq 0} q(x) \right] .$$

The next two results show the importance of the Rayleigh quotient of a self-adjoint operator.

PROPOSITION 3.39. *Let H be a Hilbert space and $T \in B(H,H)$ a self-adjoint operator. Then*

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

PROOF. Let

$$M = \sup_{\|x\|=1} |(Tx, x)| \ .$$

Obviously,

$$M \le \|T\| \ .$$

If $T \equiv 0$, we are done, so let $z \in H$ be such that $Tz \neq 0$ and $\|z\| = 1$. Set

$$v = \|Tz\|^{1/2} z \quad , \quad w = \|Tz\|^{-1/2} Tz \ .$$

Then

$$\|v\|^2 = \|w\|^2 = \|Tz\|$$

and, since T is self-adjoint

$$\Big(T(v+w), v+w\Big) - \Big(T(v-w), v-w\Big) = 2\Big[(Tv, w) + (Tw, v)\Big] = 4\|Tz\|^2 \ ,$$

and

$$\begin{aligned}\Big|\Big(T(v+w), v+w\Big) - \Big(T(v-w), v-w\Big)\Big| &\le \Big|\Big(T(v+w), v+w\Big)\Big| + \Big|\Big(T(v-w), v-w\Big)\Big| \\ &\le M \left(\|v+w\|^2 + \|v-w\|^2\right) \\ &= 2M \left(\|v\|^2 + \|w\|^2\right) \\ &= 4M\|Tz\| \ .\end{aligned}$$

We conclude that

$$\|Tz\| \le M \ ,$$

and, taking the supremum over all such z,

$$\|T\| \le M \ .$$

Thus $\|T\| = M$. □

PROPOSITION 3.40. *Let H be a Hilbert space and $T \in B(H, H)$ self-adjoint. Then*

$$r = \inf_{\|x\|=1} (Tx, x) \in \sigma(T)$$

and

$$R = \sup_{\|x\|=1} (Tx, x) \in \sigma(T) \ .$$

That is, the minimal real number in $\sigma(T)$ is r, and the maximal number in $\sigma(T)$ is R, the infimal and supremal values of the Rayleigh quotient.

PROOF. Obviously, $\lambda \in \sigma(T)$ if and only if $\lambda + \mu \in \sigma(T_\mu)$, so by such a translation, we may assume that $0 \le r \le R$. Then $\|T\| = R$ and there is a sequence $\{x_n\}_{n=1}^\infty$ such that $\|x_n\| = 1$ and

$$(Tx_n, x_n) = R - \frac{1}{n} \ .$$

Now

$$\begin{aligned}\|T_R x_n\|^2 &= \|Tx_n - Rx_n\|^2 \\ &= \|Tx_n\|^2 - 2R(Tx_n, x_n) + R^2 \\ &\le 2R^2 - 2R\Big(R - \frac{1}{n}\Big) = \frac{2R}{n} \to 0\ .\end{aligned}$$

Thus T_R is not bounded below, so $R \notin \rho(T)$, i.e., $R \in \sigma(T)$. Similar arguments show $r \in \sigma(T)$. □

We know that if $T \in B(H,H)$ is self-adjoint, then $(Tx,x) \in \mathbb{R}$ for all $x \in H$.

DEFINITION. If H is a Hilbert space and $T \in B(H,H)$ satisfies

$$(Tx,x) \ge 0 \quad \forall\, x \in H\ ,$$

then T is said to be a *positive operator*.

PROPOSITION 3.41. *Suppose H is a complex Hilbert space and $T \in B(H,H)$. Then T is a positive operator if and only if $\sigma(T) \ge 0$. Moreover, if T is positive, then T is self-adjoint.*

PROOF. This follows from Proposition 3.33 and Proposition 3.38. □

EXAMPLE. Let $H = L_2(\Omega)$ for some $\Omega \subset \mathbb{R}^d$ and $\phi : \Omega \to \mathbb{R}$ a positive and bounded function. Then $T : H \to H$ defined by

$$(Tf)(x) = \phi(x)f(x) \quad \forall\, x \in \Omega$$

is a positive operator.

An interesting and useful fact about a positive operator is that it has a square root.

DEFINITION. Let H be a Hilbert space and $T \in B(H,H)$ be positive. An operator $S \in B(H,H)$ is said to be a *square root* of T if

$$S^2 = T\ .$$

If, in addition, S is positive, then S is called a *positive square root* of T, denoted by

$$S = T^{1/2}\ .$$

THEOREM 3.42. *Every positive operator $T \in B(H,H)$, where H is a Hilbert space, has a unique positive square root.*

The proof is long but not difficult. We omit it and refer the interested reader to [**Kr**, p. 473–479].

3.8. Compact operators on a Banach space

An important class of operators exhibit a compactness property. We will see examples later.

DEFINITION. Suppose X and Y are NLS. An operator $T : X \to Y$ is a *compact linear operator* (or *completely continuous linear operator*) if T is linear and if the closure of the image of any bounded set $M \subset X$ is compact, i.e., $\overline{T(M)} \subset Y$ is compact. (We call a set with compact closure *precompact*.)

PROPOSITION 3.43. *Let X and Y be NLS. If $T : X \to Y$ is a compact linear operator, then T is bounded, hence continuous.*

Proof. The unit sphere $U = \{x \in X : \|x\| = 1\}$ in X is bounded, so $\overline{T(U)}$ is compact. A compact set in Y is necessarily bounded, so there is some $R > 0$ such that

$$\overline{T(U)} \subset B_R(0) \subset Y \ ;$$

that is,

$$\|T\| = \sup_{x \in U} \|Tx\| \le R < \infty \ ,$$

so $T \in B(X, Y)$. □

Compactness gives us convergence of subsequences, as the next two lemmas show.

Lemma 3.44. *Suppose (X, d) is a metric space. Then X is compact if and only if every sequence $\{x_n\}_{n=1}^\infty \subset X$ has a convergent subsequence $\{x_{n_k}\}_{k=1}^\infty$.*

Proof. Suppose X is compact, but that there is a sequence with no convergent subsequence. For each n, let

$$\delta_n = \inf_{m \ne n} d(x_n, x_m) \ .$$

If, for some n, $\delta_n = 0$, then there are x_{m_k} such that

$$d(x_n, x_{m_k}) < \frac{1}{k} \ ,$$

that is, $x_{m_k} \to x_n$ as $k \to \infty$, a contradiction. So $\delta_n > 0 \ \forall \ n$, and

$$\Big\{ B_{\delta_n}(x_n) \Big\}_{n=1}^\infty \cup \left(\bigcup_{n=1}^\infty \overline{B_{\delta_{n/2}}(x_n)} \right)^c$$

is an open cover of X with no finite subcover, contradicting the compactness of X and establishing the forward implication.

Suppose now that every sequence in X has a convergent subsequence. Let $\{U_\alpha\}_{\alpha \in \mathcal{I}}$ be a minimal open cover of X. By this we mean that no U_α may be removed from the collection if it is to remain a cover of X. Thus for each $\alpha \in \mathcal{I}$, $\exists \ x_\alpha \in X$ such that $x_\alpha \in U_\alpha$ but $x_\alpha \notin U_\beta$ $\forall \ \beta \ne \alpha$. If $\mathcal{I}$ is infinite, we can choose $\alpha_n \in \mathcal{I}$ for $n = 1, 2, \ldots$ and a subsequence that converges:

$$x_{\alpha_{n_k}} \to x \in X \quad \text{as} \quad k \to \infty \ .$$

Now $x \in U_\gamma$ for some $\gamma \in \mathcal{I}$. But then $\exists \ N > 0$ such that for all $k \ge N$, $x_{\alpha_{n_k}} \in U_\gamma$, a contradiction. Thus any minimal open cover is finite, and so X is compact. □

Lemma 3.45. *Let X and Y be NLS's and $T : X \to Y$ linear. Then T is compact if and only if T maps every bounded sequence $\{x_n\}_{n=1}^\infty \subset X$ onto a sequence $\{Tx_n\}_{n=1}^\infty \subset Y$ with a convergent subsequence.*

Proof. If T is compact and $\{x_n\}_{n=1}^\infty$ bounded, then the closure in Y of $\{Tx_n\}_{n=1}^\infty$ is compact. Since Y is a metric space, the conclusion follows from the previous lemma.

Conversely, suppose every bounded sequence $\{x_n\}_{n=1}^\infty$ gives rise to a convergent subsequence $\{Tx_n\}_{n=1}^\infty$. Let $B \subset X$ be bounded and consider $\overline{T(B)}$. This set is compact if every sequence $\{y_n\}_{n=1}^\infty \subset \overline{T(B)}$ has a convergent subsequence. For each $y_n \in \partial T(B)$, choose $\{y_{n,m}\}_{m=1}^\infty \subset T(B)$ such that

$$\|y_{n,m} - y_n\| \le \frac{1}{m}$$

and $x_{n,m} \in B$ such that $y_{n,m} = Tx_{n,m}$. Then $\{x_{n,n}\}_{n=1}^\infty$ is bounded and there is a convergent subsequence

$$y_{n_k,n_k} = Tx_{n_k,n_k} \to y \in \overline{T(B)} \quad \text{as} \quad k \to \infty \ .$$

But then

$$\begin{aligned}\|y_{n_k} - y\| &\le \|y_{n_k,n_k} - y\| + \|y_{n_k} - y_{n_k,n_k}\| \\ &\le \|y_{n_k,n_k} - y\| + \frac{1}{n_k} \to 0 \quad \text{as} \quad k \to \infty \ ,\end{aligned}$$

so $y_{n_k} \to y$ and, by the previous lemma, $\overline{T(B)}$ is compact. □

Trivial examples of compact operators abound, as shown by the following proposition.

PROPOSITION 3.46. *Let X and Y be NLS's and $T : X \to Y$ a linear operator. Then*

(a) *If X is finite dimensional, then T is compact.*
(b) *If T is bounded and Y is finite dimensional, then T is compact.*
(c) *If X is infinite dimensional, then $I : X \to X$ is not compact.*

PROOF. For (a), we note that necessarily T is bounded when T is linear and $\dim X < \infty$, and $R(T)$ is finite dimensional. Thus (a) follows from (b), which is trivial since closed bounded sets in finite dimensional spaces are compact. The non compactness of such sets in infinite dimensions gives (c). □

We denote the collection of all compact operators $T : X \to Y$ by

$$C(X, Y) \subset B(X, Y) \ .$$

Clearly $C(X, Y)$ is a linear subspace, as a finite linear combination of compact linear operators is compact. This set is also closed in $B(X, Y)$ when Y is complete, by the following theorem.

THEOREM 3.47. *Suppose X is a NLS and Y a Banach space. Let $\{T_n\}_{n=1}^\infty \subset C(X, Y)$ be convergent in norm to $T \in B(X, Y)$,*

$$\|T_n - T\| \to 0 \quad \textit{as} \quad n \to \infty \ .$$

Then $T \in C(X, Y)$. That is, $C(X, Y)$ is a closed linear subspace of $B(X, Y)$.

PROOF. We make extensive use of Lemma 3.45. Let $\{x_n\}_{n=1}^\infty \subset X$ be bounded. Then $\{T_1 x_n\}_{n=1}^\infty \subset Y$ has a convergent subsequence. Denote it by $\{T_1 x_{1,n}\}_{n=1}^\infty$. Then $\{x_{1,n}\}_{n=1}^\infty$ is bounded, so $\{T_2 x_{1,n}\}_{n=1}^\infty$ has a convergent subsequence. Denote it by $\{T_2 x_{2,n}\}_{n=1}^\infty$. Continuing, we obtain subsequences of $\{x_n\}_{n=1}^\infty$ satisfying

$$\{x_{k,n}\}_{n=1}^\infty \supset \{x_{k+1,n}\}_{n=1}^\infty \quad \forall \ k$$

and $T_n x_{n,m}$ converges as $m \to \infty$. We now apply a *diagonalization argument* by considering the sequence

$$\{x_{n,n}\}_{n=1}^\infty \equiv \{\tilde{x}_n\}_{n=1}^\infty \subset X \ .$$

For each $n \ge 1$, the sequence $\{T_n \tilde{x}_m\}_{m=1}^\infty$ converges, since convergence depends only on the tail of the sequence. We claim also that $\{T\tilde{x}_m\}_{m=1}^\infty$ is Cauchy, and therefore T is compact. Let $\varepsilon > 0$ be given and find $N \ge 1$ such that

$$\|T_N - T\| < \varepsilon \ .$$

Let M bound $\{x_n\}_{n=1}^\infty$. Then for any $\tilde{x}_n$ and $\tilde{x}_m$,

$$\begin{aligned}\|T\tilde{x}_n - T\tilde{x}_m\| &\le \|T\tilde{x}_n - T_N\tilde{x}_n\| + \|T_N\tilde{x}_n - T_N\tilde{x}_m\| + \|T_N\tilde{x}_m - T\tilde{x}_m\| \\ &\le 2\varepsilon M + \|T_N\tilde{x}_n - T_N\tilde{x}_m\| \ .\end{aligned}$$

Since the last term above tends to zero as $n, m \to \infty$, we have our desired conclusion. $\square$

EXAMPLE. Let $X = Y = \ell_2$ and define $T \in B(X, X)$ by

$$Tx = T(x_1, x_2, \dots) = \Big(x_1, \frac{1}{2}x_2, \frac{1}{3}x_3, \dots\Big) \ .$$

If we define

$$T_n x = \Big(x_1, \frac{1}{2}x_2, \dots, \frac{1}{n}x_n, 0, \dots\Big) \ ,$$

then T_n is compact. But

$$\|T - T_n\|^2 = \sup_{\|x\|=1} \|T_n x - Tx\|^2 = \sup_{\|x\|=1} \sum_{j=n+1}^\infty \frac{1}{j^2}|x_j|^2 \le \sum_{j=n+1}^\infty \frac{1}{j^2} \ ,$$

the tail of a convergent sequence. Thus $T_n \to T$, and we conclude that T is compact.

A useful property of a compact operator $T : X \to Y$ is that it is sequentially continuous when X has the weak topology.

THEOREM 3.48. *Suppose X and Y are NLS's and $T \in C(X, Y)$. If $\{x_n\}_{n=1}^\infty \subset X$ is weakly convergent to $x \in X$, i.e.,*

$$x_n \rightharpoonup x$$

then we have the norm or strong convergence for a subsequence

$$Tx_{n_k} \to Tx \ .$$

PROOF. Let $y_n = Tx_n$ and $y = Tx$. We first show that $y_n \rightharpoonup y$. Let $g \in Y^*$ and define $f : X \to \mathbb{F}$ by

$$f(z) = g(Tz) \ .$$

Then f is clearly linear and continuous, and so

$$f(x_n) \to f(x) \ ;$$

that is,

$$g(y_n) \to g(y)$$

and we conclude $y_n \rightharpoonup y$.

Suppose y_n does *not* converge strongly to y. Then $\{y_n\}_{n=1}^\infty$ has a subsequence $\{y_{n_k}\}_{k=1}^\infty$ such that

$$\|y - y_{n_k}\| \ge \varepsilon \quad \forall\ k \tag{3.6}$$

for some $\varepsilon > 0$. Since $\{x_n\}_{n=1}^\infty$ converges weakly, it is also bounded. Thus $\{Tx_{n_k}\}_{k=1}^\infty$ has a convergent subsequence $\{Tx_{\tilde{n}_j}\}_{j=1}^\infty$ with limit, say, $\tilde{y} \in Y$. That is, $Tx_{\tilde{n}_j} \to \tilde{y}$ as $j \to \infty$. But then also $Tx_{\tilde{n}_j} \rightharpoonup \tilde{y}$, so $\tilde{y} = y$. But

$$y_{\tilde{n}_j} = Tx_{\tilde{n}_j} \to y$$

contradicts (3.6) $\square$

PROPOSITION 3.49. *Suppose X is a NLS, $T \in C(X, X)$. Then $\sigma_p(T)$ is countable (it could be empty) and its only possible accumulation point is* 0.

PROOF. Let $r > 0$ be given. If it can be established that

$$\sigma_p(T) \cap \{\lambda : |\lambda| \geq r\}$$

is finite for any positive r, then the result follows.

Arguing by contradiction, suppose there is an $r > 0$ and a sequence $\{\lambda_n\}_{n=1}^{\infty}$ of distinct eigenvalues of T with $|\lambda_n| \geq r > 0$, for all n. Let $\{x_n\}_{n=1}^{\infty}$ be corresponding eigenvectors, $x_n \neq 0$ of course. The set $\{x_n : n = 1, 2, \dots\}$ is a linearly independent set in X, for if

$$\sum_{j=1}^{N} \alpha_j x_j = 0 \tag{3.7}$$

and N is chosen to be minimal with this property consistent with not all the α_j being zero, then

$$0 = T_{\lambda_N}\left(\sum_{j=1}^{N} \alpha_j x_j\right) = \sum_{j=1}^{N} \alpha_j(\lambda_j - \lambda_N) x_j \ .$$

Since $\lambda_j - \lambda_N \neq 0$ for $1 \leq j < N$, by the minimality of N, we conclude that $\alpha_j = 0$, $1 \leq j \leq N-1$. But then $\alpha_N = 0$ since $x_N \neq 0$. We have reached a contradiction unless (3.7) implies $\alpha_j = 0$, $1 \leq j \leq N$.

Define

$$M_n = \operatorname{span}\{x_1, \dots, x_n\} \ ,$$

and let $x \in M_n$. Then $x = \sum_{j=1}^{n} \alpha_j x_j$ for some $\alpha_j \in \mathbb{F}$. Because $Tx_j = \lambda_j x_j$, $T : M_n \to M_n$ for all n. Moreover, as above, for $x \in M_n$,

$$T_{\lambda_n} x = \sum_{j=1}^{n} \alpha_j(\lambda_j - \lambda_n) x_j = \sum_{j=1}^{n-1} \alpha_j(\lambda_j - \lambda_n) x_j \ .$$

Thus it transpires that

$$T_{\lambda_n}(M_n) \subset M_{n-1} \ , \qquad n = 1, 2, \dots \ .$$

Let $y \in M_n \setminus M_{n-1}$ and let

$$d = \operatorname{dist}\{y, M_{n-1}\} > 0 \ .$$

Then there is a $y_0 \in M_{n-1}$ such that

$$d \leq \|y - y_0\| \leq 2d \ ,$$

say. Let $z_n = (y - y_0)/\|y - y_0\|$ so that $\|z_n\| = 1$. Let $w \in M_{n-1}$ be arbitrary and note that

$$\begin{aligned} \|z_n - w\| &= \left\| \frac{1}{\|y - y_0\|}(y - y_0) - w \right\| \\ &= \frac{1}{\|y - y_0\|} \Big\| y - y_0 - \|y - y_0\| w \Big\| \\ &\geq \frac{1}{\|y - y_0\|} d \geq \frac{1}{2} \ , \end{aligned}$$

since $y_0 + \|y - y_0\| w \in M_{n-1}$.

Thus there is a sequence $\{z_n\}_{n=1}^{\infty}$ in X for which $z_n \in M_n$, $\|z_n\| = 1$ and

$$\|z_n - w\| \geq \frac{1}{2} \quad \text{for all} \quad w \in M_{n-1} \ . \tag{3.8}$$

Let $n > m$ and consider

$$Tz_n - Tz_m = \lambda_n z_n - \tilde{x}$$

where

$$\tilde{x} = \lambda_n z_n - Tz_n + Tz_m = -T_{\lambda_n} z_n + Tz_m \ .$$

As above, $T_{\lambda_n} z_n \in M_{n-1}$ and $Tz_m \in M_m \subset M_{n-1}$. Thus $\tilde{x} \in M_{n-1}$, and because of (3.8), we adduce that $(x = \tilde{x}/|\lambda_n| \in M_{n-1})$

$$\|Tz_n - Tz_m\| = |\lambda_n| \, \|z_n - x\| \geq \frac{1}{2}|\lambda_n| \geq \frac{1}{2} r > 0 \ .$$

Thus $\{Tz_n\}_{n=1}^{\infty}$ has no convergent subsequence, and this is contrary to the hypothesis that T is compact and the fact that $\{z_n\}_{n=1}^{\infty}$ is a bounded sequence. $\square$

PROPOSITION 3.50. *Suppose that X is a NLS and $T \in C(X, X)$. If $\lambda \neq 0$, then $N(T_\lambda)$ is finite dimensional.*

PROOF. If $\lambda \notin \sigma_p(T)$, then $\dim\{N(T_\lambda)\} = 0$, so we can assume $\lambda \in \sigma_p(T)$. Let B be the closed unit ball in $N(T_\lambda)$, so that

$$B = \overline{B_1(0)} \cap N(T_\lambda) \ .$$

Let $\{x_n\}_{n=1}^{\infty}$ be any sequence in B. Since B is bounded, there is a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that $\{Tx_{n_k}\}_{k=1}^{\infty}$ converges, say

$$Tx_{n_k} \to z \quad \text{as} \quad k \to \infty \ .$$

But $Tx_{n_k} = \lambda x_{n_k}$ and $\lambda \neq 0$, so $x_{n_k} \to \frac{1}{\lambda} z = w$, say. As B is closed, $w \in B$. Thus B is sequentially compact, thus compact. Since $N(T_\lambda)$ is a Hilbert space, its closed unit ball can be compact only if

$$\dim N(T_\lambda) < +\infty \ . \qquad \square$$

THEOREM 3.51. *Let X be a Banach space and $T \in C(X, X)$. If $\lambda \in \sigma(T)$ and $\lambda \neq 0$, then $\lambda \in \sigma_p(T)$. That is, all nonzero spectral values are eigenvalues.*

PROOF. Let $\lambda \in \sigma(T)$ and $\lambda \neq 0$. If $\lambda \notin \sigma_p(T)$, then T_λ is one-to-one but $R(T_\lambda) \neq X$.

Consider the nested sequence of closed subspaces

$$X \underset{\neq}{\supseteq} R(T_\lambda) \supseteq R(T_\lambda^2) \supseteq \cdots \supseteq R(T_\lambda^n) \supseteq \cdots \ .$$

This sequence must stabilize for some $n \geq 1$, which is to say

$$R(T_\lambda^n) = R(T_\lambda^{n+1}) \ .$$

If not, then use the construction in the last proposition to produce a sequence $\{x_n\}_{n=0}^{\infty}$, with

$$x_n \in R(T_\lambda^n) \ , \quad \|x_n\| = 1 \qquad n = 0, 1, \dots \ ,$$

where $R(T_\lambda^0) = R(I) = X$ by convention, having the property

$$\|x_n - x\| \geq \frac{1}{2} \quad \text{for all} \quad x \in R(T_\lambda^{n+1}) \ .$$

As before, if $n > m$, then

$$Tx_m - Tx_n = T_\lambda x_m - T_\lambda x_{m_n} + \lambda x_m - \lambda x_{m_n} = \lambda x_m - \tilde{x} \ ,$$

where

$$\tilde{x} = \lambda x_n + T_\lambda x_n - T_\lambda x_m \equiv \lambda x \ .$$

But $x_n \in R(T_\lambda^n)$, $T_\lambda x_n \in R(T_\lambda^{n+1}) \subseteq R(T_\lambda^n)$ and $T_\lambda x_m \in R(T_\lambda^{m+1})$. Hence $\tilde{x} \in R(T_\lambda^{m+1})$, and

$$\|\lambda x_m - \tilde{x}\| = |\lambda| \, \|x_m - x\| \geq \frac{1}{2}|\lambda| \ .$$

Hence $\{x_n\}_{n=1}^\infty$ is a bounded sequence such that $\{Tx_n\}_{n=1}^\infty$ has no convergent subsequence, a contradiction to the compactness of T.

Thus there is an $n \geq 1$ for which

$$R(T_\lambda^n) = R(T_\lambda^{n+1}) \ .$$

Let $y \in X \setminus R(T_\lambda)$. Consider $T_\lambda^n y \in R(T_\lambda^n) = R(T_\lambda^{n+1})$. There is an x such that

$$T_\lambda^{n+1} x = T_\lambda^n y \ ,$$

so

$$T_\lambda^n (y - T_\lambda x) = 0 \ .$$

As T_λ is one-to-one, this means

$$y - T_\lambda x = 0 \ ,$$

i.e., $y \in R(T_\lambda)$, a contradiction. □

Corollary 3.52 (Fredholm alternative). *Suppose X is a Banach space, $\lambda \in \mathbb{F}$, $\lambda \neq 0$, and $T \in C(X, X)$. Let $y \in X$ and consider*

$$(T - \lambda I)x = T_\lambda x = y \ . \tag{3.9}$$

Either

(a) *there exists a unique solution $x \in X$ to (3.9) for any $y \in X$; or*
(b) *there is some $y \in X$ with no solution, and if $y \in X$ has one solution, then it has infinitely many solutions.*

Proof. Exercise. Look at the possible spectral values of T. □

3.9. Compact self-adjoint operators on a Hilbert space

On a Hilbert space, we can be very specific about the structure of a self-adjoint, compact operator. In this case, the spectrum is real, countable, and nonzero values are eigenvalues with finite dimensional eigenspaces. Moreover, if the number of eigenvalues is infinite, then they converge to 0.

Theorem 3.53 (Hilbert-Schmidt). *Let H be a Hilbert space, $T \in C(H, H)$, and $T = T^*$. There is an* ON *set $\{u_n\}$ of eigenvectors corresponding to non-zero eigenvalues $\{\lambda_n\}$ of T such that every $x \in H$ has a unique decomposition of the form*

$$x = \sum \alpha_n u_n + v \ ,$$

where $\alpha_n \in \mathbb{C}$ and $v \in N(T)$.

PROOF. By Proposition 3.40, there is an eigenvalue λ_1 of T such that

$$|\lambda_1| = \sup_{\|x\|=1} |(Tx,x)| \ .$$

Let u_1 be an associated eigenvector, normalized so that $\|u_1\| = 1$. Let $Q_1 = \{u_1\}^\perp$. Then Q_1 is a closed linear subspace of H, so Q_1 is a Hilbert space in its own right. Moreover, if $x \in Q_1$, we have by self-adjointness that

$$(Tx, u_1) = (x, Tu_1) = \lambda_1 (x, u_1) = 0 \ ,$$

so $Tx \in Q_1$. Thus $T : Q_1 \to Q_1$ and we may conclude by Proposition 3.40 that there is an eigenvalue λ_2 with

$$|\lambda_2| = \sup_{\substack{\|x\|=1 \\ x\in Q_1}} |(Tx,x)| \ .$$

Let u_2 be a normalized eigenvector corresponding to λ_2. Plainly, $u_1 \perp u_2$. Let

$$Q_2 = \{x \in Q_1 : x \perp u_2\} = \{u_1, u_2\}^\perp \ .$$

Arguing inductively, there obtains a sequence of closed linear subspaces $\{Q_n\}$. At the n-th stage, we note that if $x \in Q_n = \{u_1, \dots , u_n\}^\perp$, then for $j = 1, \dots , n$,

$$(Tx, u_j) = (x, Tu_j) = \lambda_j (x, u_j) = 0 \ ,$$

so $T : Q_n \to Q_n$. Thus there is an eigenvalue λ_{n+1} with

$$|\lambda_{n+1}| = \sup_{\substack{\|x\|=1 \\ x\in Q_n}} |(Tx,x)|$$

and an eigenvector u_{n+1} with $\|u_{n+1}\| = 1$ corresponding to λ_{n+1}.

Two possibilities occur. Either we reach a point where $(Tx, x) > 0 \ \forall \ x \in Q_n$ but

$$(Tx, x) = 0 \tag{3.10}$$

for all $x \in Q_{n+1}$ for some n, or we don't. If (3.10) obtains, then with $T_1 = T|_{Q_{n+1}}$, our theory shows that

$$\|T_1\| = \sup_{\substack{\|x\|=1 \\ x\in Q_{n+1}}} |(Tx,x)| = 0 \ .$$

Hence T vanishes on Q_{n+1}, and $Q_{n+1} \subset N(T)$. Equality must hold since T does not vanish on $\operatorname{span}\{u_1, \dots , u_n\} \setminus \{0\}$, as $Tx = \sum_{j=1}^n \lambda_j \alpha_j u_j = 0$ only if each $\alpha_j = 0$ (the $\lambda_j \neq 0$). Thus $Q_{n+1} = N(T)$ and we have the orthogonal decomposition from $H = \operatorname{span}\{u_1, \dots , u_n\} \oplus Q_{n+1}$: Every $x \in H$ may be written uniquely as

$$x = \sum_{j=1}^n \alpha_j u_j + v$$

for some $v \in \{u_1, \dots , u_n\}^\perp = Q_{n+1}$.

If the procedure does not terminate in a finite number of steps, it generates an infinite sequence of eigenvalues $\{\lambda_n\}_{n=1}^\infty$ and eigenvectors $\{u_n\}_{n=1}^\infty$. By our general results, we know that although the λ_n may repeat, each can do so only a finite number of times. Thus

$$\lambda_n \to 0 \ \text{ as } \ n \to \infty \ .$$

Let H_1 be the Hilbert space generated by the ON family $\{u_n\}_{n=1}^{\infty}$. Every element $x \in H$ is written uniquely in the form

$$x = \sum_{j=1}^{\infty} (x, u_j) u_j + v$$

for some $v \in H_1^{\perp}$, since $H = H_1 \oplus H_1^{\perp}$. It remains to check that $H_1^{\perp} = N(T)$. Let $v \in H_1^{\perp}$, $v \neq 0$. Now,

$$H_1^{\perp} \subset Q_n \quad \text{for all} \quad n = 1, 2, \dots ,$$

so it must obtain that

$$\frac{|(Tv, v)|}{\|v\|^2} \leq \sup_{x \in Q_n} \frac{|(Tx, x)|}{\|x\|^2} = |\lambda_{n+1}| \ .$$

The right-hand side tends to zero as $n \to +\infty$, whereas the left-hand side does not depend on n. It follows that

$$(Tv, v) = 0 \quad \text{for all} \quad v \in H_1^{\perp} \ .$$

Thus $T_2 = T|_{H_1^{\perp}}$ vanishes, as

$$\|T_2\| = \sup_{\substack{\|v\|=1 \\ v \in H_1^{\perp}}} |(Tv, v)| = 0 \ ,$$

so $H_1^{\perp} \subset N(T)$. For $x \in H_1$, for some scalars β_n,

$$Tx = T\left(\sum_{n=1}^{\infty} \beta_n u_n \right) = \sum_{n=1}^{\infty} \beta_n T u_n = \sum_{n=1}^{\infty} \lambda_n \beta_n u_n \in H_1,$$

and we conclude that $T : H_1 \to H_1$ is one-to-one and onto (each $\lambda_n \neq 0$). Thus $N(T) \cap H_1 = \{0\}$, so $N(T) = H_1^{\perp}$. □

THEOREM 3.54 (Spectral Theorem for Self-Adjoint Compact Operators). *Let $T \in C(H, H)$ be a self-adjoint operator on a Hilbert space H. Then there exists an* ON *base $\{v_\alpha\}_{\alpha \in \mathcal{I}}$ for H such that each v_α is an eigenvector for T. Moreover, for every $x \in H$,*

$$Tx = \sum_{\alpha \in \mathcal{I}} \lambda_\alpha (x, v_\alpha) v_\alpha \ , \tag{3.11}$$

where λ_α is the eigenvalue corresponding to v_α.

PROOF. Let $\{u_n\}$ be the ON system constructed in the last theorem. Let H_1 be the closed subspace containing the $\{u_n\}$. Let $\{e_\beta\}_{\beta \in \mathcal{J}}$ be an ON base for $H_1^{\perp}$. Then

$$\{e_\beta\}_{\beta \in \mathcal{J}} \cup \{u_n\}$$

is an ON base for H. Moreover,

$$Te_\beta = 0 \quad \forall\, \beta \in \mathcal{J} \ ,$$

so the e_β are eigenvalues corresponding to the eigenvalue 0.

We know that for $x \in H$,

$$\sum_{n=1}^{N} (x, u_n) u_n$$

converges to x in H. Because T is continuous,

$$\sum_{n=1}^{N} \lambda_n (x, u_n) u_n = T\biggl(\sum_{n=1}^{N} (x, u_n) \biggr) u_n \to Tx \ .$$

That is, (3.11) holds since $\lambda_\alpha = 0$ for any index α corresponding to a $\beta \in \mathcal{J}$. $\square$

We have represented a self-adjoint $T \in C(H,H)$ as an infinite, diagonal matrix of its eigenvalues. It should come as no surprise that if T is a positive operators, S defined by

$$Sx = \sum_{\alpha \in \mathcal{I}} \sqrt{\lambda_\alpha}\ (x, u_\alpha) u_\alpha$$

is the positive square root of T. We leave it to the reader to verify this statement, as well as the implied fact that $S \in C(H,H)$.

PROPOSITION 3.55. *Let $S, T \in C(H,H)$ be self-adjoint operators on a Hilbert space H. Suppose $ST = TS$. Then there exists on* ON *base $\{v_\alpha\}_{\alpha \in I}$ for H of common eigenvectors of S and T.*

PROOF. Let $\lambda \in \sigma(S)$ and let V_λ be the corresponding eigenspace. For any $x \in V_\lambda$,

$$STx = TSx = T(\lambda x) = \lambda T x \Rightarrow Tx \in V_\lambda \ .$$

Therefore $T : V_\lambda \to V_\lambda$. Now T is self-adjoint on V_λ and compact, so it has a complete ON set of T-eigenvectors. This ON set are also eigenvectors for S since everything in V_λ is such. $\square$

3.10. The Ascoli-Arzela Theorem

We now discuss important examples of compact operators called integral operators. These are operators of the form

$$(Tf)(x) = \int_\Omega K(x,y) f(y)\, dy \ ,$$

where f is in an appropriate Hilbert (or Banach) space and K satisfies appropriate hypothesis. To demonstrate compactness, we will derive a more general result, known as the Ascoli-Arzela Theorem, about compact metric spaces.

LEMMA 3.56. *A compact metric space (M,d) is separable (i.e., it has a countable dense subset).*

PROOF. For any integer $n \geq 1$, cover M by balls of radius $1/n$:

$$M = \bigcup_{x \in M} B_{1/n}(x) \ .$$

By compactness, we can extract a finite subcover

$$M = \bigcup_{i=1}^{N_n} B_{1/n}(x_i^n) \tag{3.12}$$

for some $x_i^n \in M$. The set

$$S = \{x_i^n \mid i = 1, \ldots, N_n \ ; \ n = 1, 2 \ldots\}$$

is countable, and we claim that it is dense in M. Let $x \in M$ and $\varepsilon > 0$ be given. For n large enough that $1/n \le \varepsilon$, by (3.12), there is some $x_j^n \in S$ such that

$$x \in B_{1/n}(x_j^n) \ ;$$

that is, $d(x, x_j^n) < 1/n \le \varepsilon$. Thus indeed S is dense. □

THEOREM 3.57 (Ascoli-Arzela). *Let (M,d) be a compact metric space and let*

$$C(M) = C(M; \mathbb{F})$$

denote the Banach space of continuous functions from M to $\mathbb{F}$ with the maximum norm

$$\|f\| = \max_{x \in M} |f(x)| \ .$$

Let $A \subset C(M)$ be a subset that is equibounded and equicontinuous, which is to say, respectively, that

$$A \subset B_R(0)$$

for some $R > 0$ and, given $\varepsilon > 0$ there is $\delta > 0$ such that

$$\sup_{f \in A} \max_{d(x,y)<\delta} |f(x) - f(y)| < \varepsilon \ . \tag{3.13}$$

Then the closure of $A, \bar{A}$, is compact in $C(M)$.

PROOF. It suffices by Lemma 3.44 to show that an arbitrary sequence $\{f_n\}_{n=1}^\infty \subset A$ has a convergent subsequence. For each fixed $x \in M$, $\{f_n(x)\}_{n=1}^\infty$ is bounded in $\mathbb{F}$ by R, and so it has a convergent subsequence. Let $\{x_j\}_{j=1}^\infty$ be a countable dense subset of M. By a diagonalization argument, we can extract a single subsequence $\{f_{n_k}\}_{k=1}^\infty$ such that $\{f_{n_k}(x_j)\}_{k=1}^\infty$ converges for each j. The argument is as follows. Let $\{f_{n_k(x_1)}(x_1)\}_{k=1}^\infty$ be convergent, and from the bounded set $\{f_{n_k(x_1)}(x_2)\}_{k=1}^\infty$, select a convergent subsequence $\{f_{n_k(x_2)}(x_2)\}_{k=1}^\infty$. Continuing, we obtain indices

$$\{n_k(x_1)\}_{k=1}^\infty \supset \{n_k(x_2)\}_{k=1}^\infty \supset \cdots$$

such that $\{f_{n_k(x_i)}(x_j)\}_{k=1}^\infty$ converges for all $j \le i$. Finally, $\{f_{n_k(x_k)}\}_{k=1}^\infty$ is our desired subsequence.

Now let $\varepsilon > 0$ be given and fix $x \in M$. Let $\delta > 0$ correspond to ε via (3.13). There exists a finite subset $\{\tilde{x}_m\}_{m=1}^N \subset \{x_j\}_{j=1}^\infty$ such that

$$\bigcup_{m=1}^N B_\delta(\tilde{x}_m) \supset M \ ,$$

since M is compact. Choose $\tilde{x}_\ell$ such that

$$d(x, \tilde{x}_\ell) < \delta \ .$$

Then for any i, j, by (3.13),

$$\begin{aligned} &|f_{n_i}(x) - f_{n_j}(x)| \\ &\qquad \le |f_{n_i}(x) - f_{n_i}(\tilde{x}_\ell)| + |f_{n_i}(\tilde{x}_\ell) - f_{n_j}(\tilde{x}_\ell)| + |f_{n_j}(\tilde{x}_\ell) - f_{n_j}(x)| \\ &\qquad \le 2\varepsilon + |f_{n_i}(\tilde{x}_\ell) - f_{n_j}(\tilde{x}_\ell)| \\ &\qquad \le 2\varepsilon + \max_{1 \le m \le N} |f_{n_i}(\tilde{x}_m) - f_{n_j}(\tilde{x}_m)| \ . \end{aligned} \tag{3.14}$$

Since each sequence of real numbers $\{f_{n_k}(\tilde{x}_m)\}_{k=1}^\infty$ is Cauchy, we conclude that $\{f_{n_k}(x)\}_{k=1}^\infty$ is also Cauchy. Now define $f : M \to \mathbb{F}$ by

$$f(x) = \lim_{k\to\infty} f_{n_k}(x) \ .$$

This is the pointwise limit. However, since the right-hand side of (3.14) is independent of x, we conclude that in fact the convergence is uniform, i.e., the convergence is in the norm of $C(M)$. $\square$

THEOREM 3.58. *Let $\Omega \subset \mathbb{R}^d$ be bounded and open, and K continuous on $\overline{\Omega} \times \overline{\Omega}$. Let $X = C(\overline{\Omega})$ and define $T : X \to X$ by*

$$Tf(x) = \int_\Omega K(x,y)f(y)\,dy$$

(that T is well defined is easily checked). Then T is compact.

PROOF. Let $\{f_n\}_{n=1}^\infty$ be bounded in M. We must show that $\{Tf_n\}_{n=1}^\infty$ has a convergent subsequence. Since $\overline{\Omega}$ is a compact metric space, the Ascoli-Arzela theorem implies the result if the image of our sequence is equibounded and equicontinuous. The former follows since

$$\|Tf_n\|_{L_\infty(\Omega)} \le \|f_n\|_{L_\infty(\Omega)} \|K\|_{L_\infty(\Omega\times\Omega)} \int_\Omega dx$$

is bounded independently of n. For equicontinuity, we compute

$$\begin{aligned}|Tf_n(x) - Tf_n(y)| &= \Big| \int_\Omega (K(x,z) - K(y,z)) f_n(z)\,dz \Big| \\ &\le \|f_n\|_{L_\infty} \sup_{z\in\overline{\Omega}} |K(x,z) - K(y,z)| \int_\Omega dx \ .\end{aligned}$$

Since K is uniformly continuous on $\overline{\Omega} \times \overline{\Omega}$, the right-side above can be made uniformly small provided $|x - y|$ is taken small enough. $\square$

By an argument based on the density of $C(\overline{\Omega})$ in $L_2(\Omega)$, and the fact that the limit of compact operators is compact, we can extend this result to $L_2(\Omega)$. The details are left to the reader.

COROLLARY 3.59. *Let $\Omega \subset \mathbb{R}^d$ be bounded and open. Suppose $K \in L_2(\Omega \times \Omega)$ and $T : L_2(\Omega) \to L_2(\Omega)$ is defined as in the previous theorem. Then T is compact.*

3.11. Sturm Liouville Theory

Suppose $I = [a,b] \subset \mathbb{R}$, $a_j \in C^{2-j}(I)$, $j = 0, 1, 2$ and $a_0 > 0$. We consider the operator $L : C^2(I) \to C(I)$ defined by

$$(Lx)(t) = a_0(t)x''(t) + a_1(t)x'(t) + a_2(t)x(t) \ .$$

Note that L is a bounded linear operator.

THEOREM 3.60 (Picard). *Given $f \in C(I)$ and $x_0, x_1 \in \mathbb{R}$, there exists a unique solution $x \in C^2(I)$ to the initial value problem* (IVP)

$$\begin{cases} Lx = f \ , \\ x(z) = x_0 \ , \ x'(a) = x_1 \ . \end{cases} \tag{3.15}$$

Consult a text on ordinary differential equations for a proof.

COROLLARY 3.61. *The null space $N(L)$ is two dimensional.*

PROOF. We construct a basis. Solve (3.15) with $f = x_1 = 0$, $x_0 = 1$. Call this solution $z_0(t)$. Clearly $z_0 \in N(L)$. Now solve for $z_1(t)$ with $f = x_0 = 0$, $x_1 = 1$. Then any $x \in N(L)$ solves (3.15) with $x_0 = x(a)$ and $x_1 = x'(a)$, so

$$x(t) = x(a)z_0(t) + x'(a)z_1(t) \ ,$$

by uniqueness. $\square$

Thus, to solve (3.15), we cannot find L^{-1} (it does not exist). Rather, the inverse operator we desire concerns both L and the initial conditions. Ignoring these conditions for a moment, we study the structure of L within the context of an inner-product space.

DEFINITION. The *formal adjoint* of L is denoted L^* and defined by $L^* : C^2(I) \to C(I)$ where

$$\begin{aligned}(L^*x)(t) &= (\bar{a}_0 x)'' - (\bar{a}_1 x)' + \bar{a}_2 x \\ &= \bar{a}_0 x'' + (2\bar{a}_0' - \bar{a}_1)x' + (\bar{a}_0'' - \bar{a}_1' + \bar{a}_2)x \ .\end{aligned}$$

The motivation is the $L_2(I)$ inner-product. If $x, y \in C^2(I)$, then

$$\begin{aligned}(Lx, y) &= \int_a^b Lx(t)\bar{y}(t)\,dt \\ &= \int_a^b [a_0 x''\bar{y} + a_1 x'\bar{y} + a_2 x\bar{y}\,]\,dt \\ &= \int_a^b x\overline{L^*y}\,dt + [a_0 x'\bar{y} - x(a_0\bar{y}\,)' + a_1 x\bar{y}\,]_a^b \\ &= (x, L^*y) + \text{ Boundary terms.}\end{aligned}$$

DEFINITION. If $L = L^*$, we say that L is *formally self-adjoint*. If a_0, a_1, and a_2 are real-valued functions, we say that L is real.

PROPOSITION 3.62. *The real operator $L = a_0D^2 + a_1D + a_2$ is formally self-adjoint if and only if $a_0' = a_1$. In this case,*

$$Lx = (a_0x')' + a_2x = D(a_0D)x + a_2x \ ,$$

i.e.,

$$L = Da_0D + a_2 \ .$$

PROOF. Note that for a real operator,

$$L^* = a_0D^2 + (2a_0' - a_1)D + (a_0'' - a_1' + a_2) \ ,$$

so $L = L^*$ if and only if

$$\begin{aligned}a_1 &= 2a_0' - a_1 \ , \\ a_2 &= a_0'' - a_1' + a_2 \ .\end{aligned}$$

That is,

$$a_1 = a_0' \quad \text{ and } \quad a_1' = a_0'' \ ,$$

or simply the former condition. Then

$$Lx = a_0D^2x + a_0'Dx + a_2x = D(a_0Dx) + a_2x \ .$$

□

Remark. If $L = a_0D^2 + a_1D + a_2$ is real but not formally self-adjoint, we can render it so by a small adjustment using the integrating factor

$$Q(t) = \frac{1}{a_0(t)} P(t) \ ,$$
$$P(t) = \exp\left(\int_a^t \frac{a_1(\tau)}{a_0(\tau)}\, d\tau\right) > 0 \ ,$$

for which $P' = a_1P/a_0$. Then

$$Lx = f \iff \tilde{L}x = \tilde{f} \ ,$$

where

$$\tilde{L} = QL \quad \text{and} \quad \tilde{f} = Qf \ .$$

But $\tilde{L}$ is formally self-adjoint, since

$$\begin{aligned} \tilde{L}x = QLx &= Px'' + \frac{a_1}{a_0} Px' + a_2Qx \\ &= Px'' + P'x' + a_2Qx \\ &= (Px') + \left(\frac{a_2}{a_0} P\right) x \ . \end{aligned}$$

Examples. The most important examples are posed for $I = (a, b)$, a or b possibly infinite, and $a_j \in C^{2-j}(\bar{I})$, where $a_0 > 0$ on I (thus $a_0(a)$ and $a_0(b)$ may vanish — we have excluded this case, but the theory is similar).

(a) Legendre:

$$Lx = ((1 - t^2)x')' \ , \qquad -1 \le t \le 1 \ .$$

(b) Chebyshev:

$$Lx = (1 - t^2)^{1/2}((1 - t^2)^{1/2}x')' \ , \qquad -1 \le t \le 1 \ .$$

(c) Laguerre:

$$Lx = e^t(te^{-t}x')' \ , \qquad 0 < t < \infty \ .$$

(d) Bessell: for $\nu \in \mathbb{R}$,

$$Lx = \frac{1}{t}(tx')' - \frac{\nu^2}{t^2}\, x \ , \qquad 0 < t < 1 \ .$$

(e) Hermite:

$$Lx = e^{t^2}(e^{-t^2}x')' \ , \qquad t \in \mathbb{R} \ .$$

We now include and generalize the initial conditions, which characterize $N(L)$. Instead of two conditions at $t = a$, we consider one condition at each end of $I = [a, b]$, called *boundary conditions* (BC's).

Definition. Let p, q, and w be real-valued functions on $I = [a, b]$, $a < b$ both finite, with $p \neq 0$ and $w > 0$. Let $\alpha_1, \alpha_2, \beta_1$, and $\beta_2 \in \mathbb{R}$ be such that

$$\alpha_1^2 + \alpha_2^2 \neq 0 \quad \text{and} \quad \beta_1^2 + \beta_2^2 \neq 0 \ .$$

Then the problem of finding $x(t) \in C^2(I)$ and $\lambda \in \mathbb{C}$ such that

$$\begin{cases} Ax \equiv \frac{1}{w}[(px')' + qx] = \lambda x \ , \qquad t \in (a,b) \ , \\ \alpha_1 x(a) + \alpha_2 x'(a) = 0 \ , \\ \beta_1 x(b) + \beta_2 x'(b) = 0 \ , \end{cases} \tag{3.16}$$

is called a *regular Sturm-Liouville* (regular SL) problem. It is the eigenvalue problem for A with the BC's.

We remark that if a or b are infinite or p vanishes at a or b, the corresponding BC is lost and the problem is called a singular SL problem .

Example. Let $I = [0,1]$ and

$$\begin{cases} Ax = -x'' = \lambda x \ , \qquad t \in (0,1) \ , \\ x(0) = x(1) = 0 \ . \end{cases} \tag{3.17}$$

Then we need to solve

$$x'' + \lambda x = 0 \ ,$$

which as we saw has the 2 dimensional form

$$x(t) = A \sin \sqrt{\lambda}\, t + b \cos \sqrt{\lambda}\, t$$

for some constants A and B. Now the BC's imply that

$$x(0) = B = 0 \ ,$$
$$x(1) = A \sin \sqrt{\lambda} = 0 \ .$$

Thus either $A = 0$ or, for some integer n,

$$\sqrt{\lambda} = n\pi \ ;$$

that is, non trivial solutions are given only for the eigenvalues

$$\lambda_n = n^2 \pi^2 \ ,$$

and the corresponding eigenfunctions are

$$x_n(t) = \sin(n\pi t)$$

(or any nonzero multiple).

To analyze a regular SL problem, it is helpful to notice that

$$A : C^2(I) \to C^0(I)$$

has strictly larger range. However, its inverse (with the BC's), would map $C^0(I)$ to $C^2(I) \subset C^0(I)$. So the inverse might be a bounded linear operator with known spectral properties, which can then be related to A itself. This is the case, and leads us to the classical notion of a Green's function. The Green's function allows us to construct the solution to the boundary value problem

$$\begin{cases} Ax = f \ , \qquad t \in (a,b) \ , \\ \alpha_1 x(a) + \alpha_2 x'(a) = 0 \ , \\ \beta_1 x(b) + \beta_2 x'(b) = 0 \ , \end{cases} \tag{3.18}$$

for any $f \in C^0(I)$.

Definition. A Green's function for the regular SL problem (3.16) is a function $G : I\times I \to \mathbb{R}$ such that

(a) $G \in C^0(I \times I)$ and $G \in C^2(I \times I \setminus D)$, where $D = \{(t,t) : t \in I\}$ is the diagonal in $I \times I$;
(b) For each fixed $s \in I$, $G(\cdot, s)$ satisfies the BC's of the problem;
(c) A applied to the first variable t of $G(t,s)$, also denoted $A_tG(t,s)$, vanishes for $(t,s) \in I \times I \setminus D$, i.e.,

$$\begin{aligned} A_tG(t,s) &\equiv \frac{1}{w}\left[\frac{\partial}{\partial t}\Big(p(t)\frac{\partial G}{\partial t}(t,s)\Big) + q(t)G(t,s)\right] \\ &= 0 \quad \forall \quad t \neq s \ ; \end{aligned}$$

(d) $\displaystyle\lim_{s\to t^-}\frac{\partial G}{\partial t}(t,s) - \lim_{s\to t^+}\frac{\partial G}{\partial t}(t,s) = \frac{1}{p(t)}$ for all $t \in (a,b)$.

Example. Corresponding to (3.17), consider

$$\begin{cases} Ax = -x'' = f \ , & t \in (0,1) \ , \\ x(0) = x(1) = 0 \ , \end{cases} \tag{3.19}$$

for $f \in C^0(I)$. Let

$$G(t,s) = \begin{cases} (1-t)s \ , & 0 \le s \le t \le 1 \ , \\ (1-s)t \ , & 0 \le t \le s \le 1 \ . \end{cases}$$

Then G satisfies (a) and

$$\begin{aligned} G(0,s) &= (1-s)\cdot 0 = 0 \ , \\ G(1,s) &= (1-(1))s = 0 \ , \end{aligned}$$

so (b) holds. Since $w = 1$, $p = -1$, and $q = 0$,

$$A_tG(t,s) = -\frac{\partial^2}{\partial t^2}G(t,s) = 0 \quad \text{for} \quad s \neq t$$

and

$$\lim_{s\to t^-}\frac{\partial G}{\partial t} = -t \quad , \quad \lim_{s\to t^+}\frac{\partial G}{\partial t} = 1-t \ ,$$

we also have (c) and (d). Thus $G(t,s)$ is our Green's function. Moreover, if we define

$$x(t) = \int_0^1 G(t,s)f(s)\,ds \ ,$$

then $x(0) = x(1) = 0$ and

$$\begin{aligned}
x'(t) &= \frac{d}{dt}\left\{\int_0^t G(t,s)f(s)\,ds + \int_t^1 G(t,s)f(s)\,ds\right\} \\
&= G(t,t)f(t) + \int_0^t \frac{\partial G}{\partial t}(t,s)f(s)\,ds - G(t,t)f(t) + \int_t^1 \frac{\partial G}{\partial t}(t,s)f(s)\,ds \\
&= \int_0^1 \frac{\partial G}{\partial t}(t,s)f(s)\,ds\ , \\
s''(t) &= \frac{d}{dt}\left\{\int_0^t \frac{\partial G}{\partial t} f\,ds + \int_t^1 \frac{\partial G}{\partial t} f\,ds\right\} \\
&= \frac{\partial G}{\partial t}(t,t^-)f(t) + \int_0^t \frac{\partial^2 G}{\partial t^2} f\,ds - \frac{\partial G}{\partial t}(t,t^+)f(t) + \int_t^1 \frac{\partial^2 G}{\partial t^2} f\,ds \\
&= -f + \int_0^1 \frac{\partial^2 G}{\partial t^2} f\,ds \\
&= -f(t)\ .
\end{aligned}$$

Thus we constructed a solution to (3.19) with $G(t,s)$.

THEOREM 3.63. *Suppose that for the regular SL system*

$$\begin{cases}
Au \equiv \dfrac{1}{w} Lu \equiv \dfrac{1}{w}[(pu')' + qu]\ , \qquad t \in (a,b)\ , \\
\alpha_1 u(a) + \alpha_2 u'(a) = 0\ , \\
\beta_1 u(b) + \beta_2 u'(b) = 0\ ,
\end{cases}$$

on the interval $I = [a,b]$, $p \in C^1(I)$, $w, q \in C^0(I)$, and $p, w > 0$. Suppose also that 0 is not an eigenvalue (so $Au = 0$ with the BC's implies $u = 0$). Let u_1 and u_2 be any nonzero real solutions of $Au = Lu = 0$ such that for u_1,

$$\alpha_1 u_1(a) + \alpha_2 u_1'(a) = 0\ ,$$

and for u_2,

$$\beta_1 u_2(b) + \beta_2 u_2'(b) = 0\ .$$

Define $G : I \times I \to \mathbb{R}$ by

$$G(t,s) = \begin{cases}
\dfrac{u_2(t)u_1(s)}{pW}\ , \qquad a \le s \le t \le b\ , \\
\dfrac{u_1(t)u_2(s)}{pW}\ , \qquad a \le t \le s \le b\ .
\end{cases}$$

where $p(t)w(t)$ is a nonzero constant and

$$W(s) = W(s; u_1, u_2) \equiv u_1(s)u' - 2(s) - u_1'(s)u_2(s)$$

is the Wronskian *of u_1 and u_2. Then G is a Green's function for L. Moreover, if $\mathcal{G}$ is any Green's function for L and $f \in C^0(I)$, then*

$$u(t) = \int_0^b \mathcal{G}(t,s)f(s)\,ds \tag{3.20}$$

is the unique solution of $Lu = f$ satisfying the BC's.

To solve $Au = f$, just solve $Lu = wf$:

$$u(t) = \int_a^b G(t,s) f(s) w(s)\, ds \ .$$

We first prove two lemmas concerning the Wronskian.

LEMMA 3.64 (Abel). *Let $Lu = (pu') + qu$ satisfy $p \in C^1(I)$ and $q \in C^0(I)$. For any positive $w \in C^0(I)$ and $\lambda \in \mathbb{C}$, if u_1 and u_2 solve*

$$Lu = \lambda w u \ ,$$

then

$$p(t) W(t; u_1, u_2)$$

is constant.

PROOF. We compute

$$\begin{aligned} 0 &= \lambda w(u_1 u_2 - u_2 u_1) \\ &= u_1 L u_2 - u_2 L u_1 \\ &= u_1(pu_2'' + p'u_2' + qu_2) - u_2(pu_1'' + p'u_1' + qu') \\ &= p(u_1 u_2'' - u_2 u_1'') + p'W \\ &= (pW)' \ . \end{aligned}$$

□

LEMMA 3.65. *Suppose $u, v \in C^1(I)$. If $W(t_0; u, v) \neq 0$ for some $t_0 \in I$, then u and v are linearly independent. If u and v are linearly independent, then $W(t; u, v) \neq 0$ for all $t \in I$.*

PROOF. Suppose for some scalars α and β,

$$\alpha u(t) + \beta v(t) = 0 \ ,$$

so also

$$\alpha u'(t) + \beta v'(t) = 0 \ .$$

At $t = t_0$, we have a linear system

$$\begin{bmatrix} u(t_0) & v(t_0) \\ u'(t_0) & v'(t_0) \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \ ,$$

which is uniquely solvable if the matrix is invertible, i.e., if its determinant, $W(t_0) \neq 0$. Thus $\alpha = \beta = 0$ and we conclude that u and v are linearly independent.

Conversely, the linear independence of u and v requires the determinant $W(t) \neq 0$ for each $t \in I$. □

PROOF OF THEOREM 3.63. The existence of u_1 and u_2 follows from Picard's Theorem 3.60. If we use the standard basis

$$N(L) = \operatorname{span}\{z_0, z_1\} \ ,$$

where

$$\begin{aligned} z_0(a) &= 1 \ , \qquad & z_0'(a) &= 0 \ , \\ z_1(a) &= 0 \ , \qquad & z_1'(a) &= 1 \ , \end{aligned}$$

then

$$u_1(t) = -\alpha_2 z_0(t) + \alpha_1 z_1(t) \not\equiv 0 \ .$$

A similar construction at $t = b$ gives $u_2(t)$.

If $u_1 = \lambda u_2$ for some $\lambda \in \mathbb{C}$, i.e., u_1 and u_2 are linearly dependent, then $u_1 \not\equiv 0$ satisfies both boundary conditions, since λ cannot vanish, and the equation $Lu_1 = 0$, contrary to the hypothesis that 0 is not an eigenvalue to the SL problem. Thus u_1 and u_2 are linearly independent, and by our two lemmas pW is a nonzero constant. Thus $G(t,s)$ is well defined.

Clearly G is continuous and C^2 when $t \neq s$, since $u_1, u_2 \in C^2(I)$. Moreover, $G(\cdot, s)$ satisfies the BC's by construction, and $A_t G$ is either $Au_1 = 0$ or $Au_2 = 0$ for $t \neq s$. Thus it remains only to show the jump condition on $\partial G/\partial t$ of the definition of a Green's function. But

$$\frac{\partial G}{\partial t}(t,s) = \begin{cases} \dfrac{u_2'(t)u_1(s)}{pW} \ , & a \le s \le t \le b \ , \\[2ex] \dfrac{u_1'(t)u_2(s)}{pW} \ , & a \le t \le s \le b \ , \end{cases}$$

so

$$\frac{\partial G}{\partial t}(s^+, s) - \frac{\partial G}{\partial t}(s^-, s) = \frac{u_2'(s)u_1(s)}{pW} - \frac{u_1'(s)u_2(s)}{pW} = \frac{1}{p(s)} \ .$$

If $Lu = f$ has a solution, it must be unique since the difference of two such solutions would satisfy the eigenvalue problem with eigenvalue 0, and therefore vanish. Thus it remains only to show that $u(t)$ defined by (3.20) is a solution to $Lu = f$. We use only (a)–(d) in the definition of a Green's function.

Trivially u satisfies the two BC's by (b) and the next computation. We compute for $t \in (a,b)$ using (a):

$$\begin{aligned} u'(t) &= \frac{d}{dt}\left(\int_a^b \mathcal{G}(t,s) f(s)\, ds \right) \\ &= \frac{d}{dt}\left(\int_a^t \mathcal{G}(t,s) f(s)\, ds \right) + \frac{d}{dt}\left(\int_t^b \mathcal{G}(t,s) f(s)\, ds \right) \\ &= \mathcal{G}(t,t) f(t) + \int_a^t \frac{\partial \mathcal{G}}{\partial t}(t,s) f(s)\, ds - \mathcal{G}(t,t) f(t) + \int_t^b \frac{\partial \mathcal{G}}{\partial t}(t,s) f(s)\, ds \\ &= \int_a^b \frac{\partial \mathcal{G}}{\partial t}(t,s) f(s)\, ds \ . \end{aligned}$$

Then

$$\begin{aligned} (p(t)u'(t))' &= \frac{d}{dt}\left(\int_a^t p(t) \frac{\partial \mathcal{G}}{\partial t}(t,s) f(s)\, ds \right) + \frac{d}{dt}\left(\int_t^b p(t) \frac{\partial \mathcal{G}}{\partial t}(t,s) f(s)\, ds \right) \\ &= p(t) \frac{\partial \mathcal{G}}{\partial t}(t,t^-) f(t) + \int_0^t \frac{\partial}{\partial t}\Big(p(t) \frac{\partial \mathcal{G}}{\partial t}(t,s) \Big) f(s)\, ds \\ &\qquad - p(t) \frac{\partial \mathcal{G}}{\partial t}(t,t^+) f(t) + \int_t^b \frac{\partial}{\partial t}\Big(p(t) \frac{\partial \mathcal{G}}{\partial t}(t,s) \Big) f(s)\, ds \\ &= f(t) + \int_a^b \frac{\partial}{\partial t}\Big(p(t) \frac{\partial \mathcal{G}}{\partial t}(t,s) \Big) f(s)\, ds \ , \end{aligned}$$

using (d). Finally, we use (c) to conclude

$$\begin{aligned}
Lu(t) &= (pu')' + qu \\
&= f(t) + \int_a^b A_t \mathcal{G}(t,s) f9s) w(t)\, ds \\
&= f(t)
\end{aligned}$$

as required. □

We define the solution operator

$$T : C^0(I) \to C^0(I)$$

by

$$Tf(t) = \int_a^b G(t,s) f(s)\, ds \ ,$$

where G is our Green's function. Endowing T with the $L_2(I)$ innerproduct, we conclude that T is a bounded linear operator, since for $f \in C^0(I)$,

$$\begin{aligned}
\|Tf\|^2 &\le \int_a^b \left(\int_a^b |G(t,s)|\, f(s)|\, ds \right)^2 dt \\
&\le \int_a^b \int_a^b |G(t,s)|^2\, ds \int_a^b |f(s)|^2\, ds\, dt \\
&= \|G\|^L_{L_2(I\times I)} \|f\|^2_{L_2(I)} \ .
\end{aligned}$$

Since $G(s,t) = G(t,s)$ is real, we compute that for $f, g \in C^0(I)$,

$$\begin{aligned}
(Tf, g) &= \int_a^b \int_a^b G(t,s) f(s)\, ds\, \overline{g(t)}\, dt \\
&= \int_a^b f(s) \int_a^b G(s,t)\, \overline{g(t)}\, dt\, ds \\
&= (f, Tg) \ ,
\end{aligned}$$

that is, T is self-adjoint. By the Ascoli-Arzela theorem, we know that T is a compact operator. The incompleteness of $C^0(I)$ is easily rectified, since $C^0(I)$ is dense in $L_2(I)$. We extend T to $L_2(I)$ as follows. Given $u \in L_2(I)$, find $u_n \in C^0(I)$ such that $u_n \to u$ in $L_2(I)$. Then boundedness implies that $\{Tu_n\}_{n=1}^\infty$ is Cauchy in $L_2(I)$. So define

$$Tu = \lim_{n\to\infty} Tu_n \ .$$

Then

$$T : L_2(I) \to L_2(I)$$

is a continuous linear operator. Moreover, it is not difficult to conclude that the extended T remains compact and self-adjoint.

We know much about the spectral properties of T. We relate these properties to those of $L = wA$.

Proposition 3.66. *If $\lambda = 0$ is not an eigenvalue of the regular SL problem, then $\lambda = 0$ is not an eigenvalue of T either.*

PROOF. Suppose $Tf = 0$ for some $f \in L_2(I)$. Then, with $c = (pW)^{-1}$,

$$\begin{aligned} 0 = (Tf)'(t) &= \frac{d}{dt}\left\{ cu_2(t)\int_a^t f(s)u_1(s)\,ds + cu_1(t)\int_t^b f(s)u_2(s)\,ds \right\} \\ &= c\left\{ u_2'\int_a^t fu_1\,ds + u_1'\int_t^b fu_2\,ds \right\} . \end{aligned}$$

But

$$0 = Tf(t) = c\left\{ u_2\int_a^t fu_1\,ds + u_1\int_t^b fu_2\,ds \right\} ,$$

so, since $W(t; u_1, u_2) \neq 0$, the solution of this linear system is trivial; that is, for each $t \in [a, b]$,

$$\int_a^t fu_1\,ds = \int_t^b fu_2\,ds = 0 .$$

We conclude that

$$f(t)u_1(t) = f(t)u_2(t) = 0 ,$$

so $f = 0$, since u_1 and u_2 cannot both vanish at the same point ($W \neq 0$). Thus $N(T) = \{0\}$ and $0 \notin \sigma_p(T)$. □

PROPOSITION 3.67. *Suppose $\lambda \neq 0$. Then λ is an eigenvalue of the regular SL problem if and only if $1/\lambda$ is an eigenvalue of T. Moreover, the corresponding eigenspaces coincide.*

PROOF. If $f \in C^0(I)$ is an eigenfunction for L, then

$$Lf = \lambda f ,$$

so

$$f = TLf = \lambda Tf$$

shows that

$$Tf = \frac{1}{\lambda} f .$$

Conversely, suppose $f \in L_2(I)$ is an eigenfunction for T:

$$Tf = \frac{1}{\lambda} f .$$

Since G is continuous, in fact $R(T) \subset C^0(I)$, so $f \in C^0(I)$ and

$$f = LTf = \frac{1}{\lambda} Lf .$$

□

We return to our original operator $A = \frac{1}{w}L$. Define the innerproduct on $L_2(I)$

$$\langle f, g\rangle_w = \int_a^b f(t)\,\overline{g(t)}\,w(t)\,dt .$$

This induces a norm equivalent to the usual $L_2(I)$-norm, since

$$0 < \min_{s\in I} w(s) \leq w(t) \leq \max_{s\in I} w(s) < \infty$$

for all $t \in I$. Define $K : L_2(I) \to L_2(I)$ by

$$Kf(t) = \int_a^b G(t,s)f(s)w(s)\,ds \ .$$

This is the solution operator for

$$Au = f \ .$$

With the usual innerproduct on $L_2(I)$, K is not self-adjoint; however, with $\langle \cdot, \cdot \rangle_w$, K is self-adjoint. The proof of the following result is left as an exercise .

PROPOSITION 3.68. *The operator K is self-adjoint and compact on $(L_2(I), \langle \cdot, \cdot \rangle_w)$, $0 \notin \sigma_p(K)$, and*

$$\sigma(K) = \{0\} \cup \{\lambda \neq 0 : 1/\lambda \text{ is an eigenvalue of } A\} \ .$$

Moreover, the eigenspaces of K and A coincide.

We know that $\dim(N(T_\lambda)) = \dim(N(K_\lambda))$ is finite. However, we can conclude directly that eigenfunctions of a regular SL problem are simple (i.e., one dimensional).

PROPOSITION 3.69. *The eigenvalues of a regular SL problem are simple.*

PROOF. Suppose u and v are eigenvectors for $\lambda \neq 0$ an eigenvalues. Lemma 3.64 tells us that $pW = c$ for some constant c. If $c = 0$, then as $p \neq 0$, $W \equiv 0$ and u and v are linearly independent. So suppose $W(t_0) \neq 0$ for some t_0. By Lemma 3.65, $W \neq 0$ for all $t \in [a, b]$. However, $W(a) = 0$ by the boundary conditions:

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \ ,$$
$$\alpha_1 v(a) + \alpha_2 v'(a) = 0 \ ,$$

is a linear system with a nontrivial solution (α_1, α_2), so $W(a)$, the determinant of the corresponding matrix, vanishes. Thus u and v are linearly independent and λ is simple. □

We summarize what we know about the regular SL problem for A based on the Spectral Theorem for Compact Self-adjoint operators as applied to K. The details of the proof are left as an exercise.

THEOREM 3.70. *Let $a, b \in \mathbb{R}$, $a < b$, $I = [a, b]$, $p \in C^1(I)$, $p \neq 0$, $q \in C^0(I)$, and $w \in C^0(I)$, $w > 0$. Let*

$$A = \frac{1}{w}[DpD + q]$$

be a formally self-adjoint regular SL operator with boundary conditions

$$\alpha_1 u(a) + \alpha_2 u'(a) = 0 \ ,$$
$$\beta_1 u(b) + \beta_2 u'(b) = 0 \ ,$$

for $u \in C^2(I)$, where $\alpha_1^2 + \alpha_2^2 \neq 0$ and $\beta_1^2 + \beta_2^2 \neq 0$, $\alpha_i, \beta_i \in \mathbb{R}$. If 0 is not an eigenvalue of A, then A has a countable collection of real eigenvalues $\{\lambda_n\}_{n=1}^\infty$ such that

$$|\lambda_n| \to \infty \quad \text{as} \quad n \to \infty$$

and each eigenspace is one-dimensional. Let $\{u_n\}_{n=1}^\infty$ *be the corresponding normalized eigenfunctions. These form an ON basis for* $(L_2(I), \langle\cdot,\cdot\rangle_w)$*, so if* $u \in L_2(I)$*,*

$$u = \sum_{n=1}^\infty \langle u, u_n\rangle_w u_n$$

and, provided $Au \in L_2(I)$*,*

$$Au = \sum_{n=1}^\infty \lambda_n \langle u, u_n\rangle_w u_n \ .$$

We saw earlier that the regular SL problem

$$\begin{cases} -x'' = \lambda x \ , & t \in (0,1) \\ x(0) = x(1) = 0 \end{cases}$$

has eigenvalues

$$\lambda_n = n\pi \ , \qquad n = 1, 2, \dots$$

and corresponding (normalized) eigenfunctions

$$u_n(t) = \sqrt{2}\sin(n\pi t) \ .$$

Given any $f \in L_2(0,1)$, we have its *sine series*

$$f(t) = \sum_{n=1}^\infty \sqrt{2}\int_0^1 f(s)\sin n\pi s\, ds \ \sin n\pi t \ ,$$

where equality holds for a.e. $t \in [0,1]$, i.e., in $L_2(0,1)$. This shows that $L_2(0,1)$ is separable.

By iterating our result, we can decompose any $f \in L_2(I\times I)$, $I = (0,1)$. For a.e. $x \in I$,

$$\begin{aligned} f(x,y) &= \sum_{n=1}^\infty \sqrt{2}\int_0^1 f(x,t)\sin n\pi t\, dt \ \sin n\pi y \\ &= 2\sum_{n=1}^\infty \int_0^1 \sum_{m=1}^\infty \int_0^1 f(s,t)\sin m\pi s\, ds \ \sin n\pi t\, dt \ \sin n\pi y \ \sin n\pi x \\ &= 2\sum_{n=1}^\infty\sum_{m=1}^\infty \int_0^1\int_0^1 f(s,t)\sin m\pi s \ \sin n\pi t\, ds\, dt \ \sin n\pi x \ \sin n\pi y \ . \end{aligned}$$

So $L_2(I\times I)$ has the ON basis

$$\{2\,\sin n\pi x\,\sin n\pi y\}_{m=1,\,n=1}^{\infty,\infty} \ ,$$

and again $L_2(I\times I)$ is separable. Continuing, we can find a countable basis for any $L_2(R)$, $R = I^d$, $d = 1, 2, \dots$. By dilation and translation, we can replace R by any rectangle, and since $L_2(\Omega) \subset L_2(R)$ whenever $\Omega \subset R$ (if we extend the domain of $f \in L_2(\Omega)$ by defining $f \equiv 0$ on $R\setminus\Omega$), $L_2(\Omega)$ is separable for any bounded Ω, but the construction of a basis is not so clear.

Example. Let $\Omega = (0,a)\times(0,b)$, and consider a solution $u(x,y)$ of

$$\begin{cases} -\dfrac{\partial^2 u}{\partial x^2} - \dfrac{\partial^2 u}{\partial y^2} = f(x,y) \ , \quad (x,y) \in \Omega \ , \\ u(x,y) = 0 \ , \quad (x,y) \in \partial\Omega \ , \end{cases}$$

where $f \in L_2(\Omega)$. We proceed formally; that is, we compute without justifying our steps. We justify the final result only. We use the technique of separation of variables. Suppose $v(x,y) = X(x)Y(y)$ is a solution to the eigenvalue problem

$$-X''Y - XY'' = \lambda XY \ .$$

Then

$$-\frac{X''}{X} = \lambda + \frac{Y''}{Y} = \mu \ ,$$

a constant. Now the BC's are

$$\begin{aligned} X(0) = X(a) = 0 \ , \\ Y(0) = Y(b) = 0 \ , \end{aligned}$$

so X satisfies a SL problem with

$$\begin{aligned} \mu = \mu_m &= \left(\frac{m\pi}{a}\right)^2 \ , \qquad m = 1, 2, \dots \ , \\ X_m(x) &= \sin\left(\frac{m\pi x}{a}\right) \ . \end{aligned}$$

Now, for each such m,

$$-Y'' = (\lambda_m - \mu_m)Y$$

has solution

$$\begin{aligned} \lambda_{m,n} - \mu_m &= \left(\frac{n\pi}{b}\right)^2 \ , \qquad n = 1, 2, \dots \ , \\ Y_n(y) &= \sin\left(\frac{n\pi y}{b}\right) \ . \end{aligned}$$

That is, for $m, n = 1, 2, \dots,$

$$\begin{aligned} \lambda_{m,n} &= \left[\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2\right]\pi^2 \ , \\ V_{m,n}(x,y) &= \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b} \ . \end{aligned}$$

We know that $\{v_{m,n}\}$ form a basis for $L_2((0,a) \times (0,b))$, so, rigorously, we expand

$$f(x,y) = \sum_{m,n} c_{m,n} v_{m,n}(x,y)$$

for the coefficients

$$c_{m,n} = \frac{\int_0^b \int_0^a f(x,y) v_{m,n}(x,y)\,dx\,dy}{\int_0^b \int_0^a v_{m,n}^2(x,y)\,dx\,dy} \ .$$

Forming

$$u(x,y) \equiv \sum_{m,n} \frac{c_{m,n}}{\sqrt{\lambda_{m,n}}} v_{m,n}(x,y) \ ,$$

we verify that indeed u is a solution to the problem.

3.12. Exercises

1. Prove the parallelogram law in a Hilbert space.

2. On a NLS X, a linear map $P : X \to X$ is a projection if $P^2 = P$.

   (a) Prove that every projection on a Hilbert space for which $\|P\| = 1$ is the orthogonal projection onto some subspace of H.

   (b) Prove that in general if $P \not\equiv 0$, $\|P\| \geq 1$. Show by example that if the Hilbert space H has at least two dimensions, then there is a nonorthogonal projection defined on H.

3. Let H be a Hilbert space, and $R : H \to H^*$ the Reisz map.

   (a) Show that R is conjugate linear.

   (b) Show that the map $(\cdot,\cdot)_{H^*} : H^* \times H^* \to \mathbb{F}$ defined by $(L_1, L_2)_{H^*} = (R^{-1}L_2, R^{-1}L_1)_H$ is an inner product.

4. Show that if $\mathcal{I}$ is an index set and $\{x_\alpha\}_{\alpha\in\mathcal{I}}$ is a collection of nonnegative real numbers satisfying
$$\sum_{\alpha\in\mathcal{I}} x_\alpha < \infty \ ,$$
then at most countably many of the x_α are different from zero.

5. If $\{u_\alpha\}_{\alpha\in\mathcal{I}}$ is a maximal ON set in a Hilbert space $(H, (\cdot,\cdot))$, and $x \in H$, show that there exist at most countably many $\alpha_i \in \mathcal{I}$ such that
$$x = \sum_{i=1}^{\infty} (x, u_{\alpha_i}) u_{\alpha_i} \ .$$

6. Prove that for any index set $\mathcal{I}$, the space $\ell_2(\mathcal{I})$ is a Hilbert space.

7. Let H be a Hilbert space and $\{x_n\}_{n=1}^\infty$ a bounded sequence in H.

   (a) Show that $\{x_n\}_{n=1}^\infty$ has a weakly convergent subsequence.

   (b) Suppose that $x_n \xrightarrow{w} x$. Prove that $x_n \to x$ if and only if $\|x_n\| \to \|x\|$.

   (c) If $x_n \xrightarrow{w} x$, then there exist non-negative constants $\{\{\alpha_i^n\}_{i=1}^n\}_{n=1}^\infty$ such that $\displaystyle\sum_{i=1}^{n} \alpha_i^n = 1$ and
$$\sum_{i=1}^{n} \alpha_i^n x_i \equiv y_n \to x \quad \text{(strong convergence)}.$$

8. Let $\{x_n\}_{n=1}^\infty$ be an orthonormal set in a Hilbert space H. Let $\{a_n\}_{n=1}^\infty$ be a sequence of non-negative numbers and let
$$S = \left\{ x \in H : x = \sum_{n=1}^{\infty} b_n x_n \text{ and } |b_n| \leq a_n \text{ for all } n \right\} .$$
Show that S is compact if and only if $\sum_{n=1}^\infty a_n^2 < \infty$.

9. Let H be a Hilbert space and Y a subspace (not necessarily closed).

   (a) Prove that
$$(Y^\perp)^\perp = \bar{Y} \qquad \text{and} \qquad Y^\perp = (\bar{Y})^\perp \ .$$

(b) If Y is not trivial, show that P, projection onto $\bar{Y}$, has norm 1 and that

$$(Px, y) = (x, y)$$

for all $x \in H$ and $y \in \bar{Y}$.

10. Let H be a Hilbert space and $P \in B(H, H)$ a projection.

(a) Show that P is an orthogonal projection if and only if $P = P^*$.

(b) If P is an orthogonal projection, find $\sigma_p(P)$, $\sigma_c(P)$, and $\sigma_r(P)$.

11. Let A be a self-adjoint, compact operator on a Hilbert space. Prove that there are positive operators P and N such that $A = P - N$ and $PN = 0$. (An operator T is positive if $(Tx, x) \geq 0$ for all $x \in H$.) Prove the conclusion if A is merely self-adjoint.

12. Let T be a compact, positive operator on a complex Hilbert space H. Show that there is a unique positive operator S on H such that $S^2 = T$. Moreover, show that S is compact.

13. Give an example of a self-adjoint operator on a Hilbert space that has no eigenvalues (see [**Kr**], p. 464, no. 9).

14. Let H be a separable Hilbert space and T a positive operator on H. Let $\{e_n\}_{n=1}^{\infty}$ be an orthonormal base for H and suppose that $tr(T)$ is finite, where

$$tr(T) = \sum_{n=1}^{\infty} (Te_n, e_n) \; .$$

Show the same is true for any other orthonormal base, and that the sum is independent of which base is chosen. Show that this is not necessarily true if we omit the assumption that T is positive.

15. Let H be a Hilbert space and $S \in B(H, H)$. Define $|S|$ to be the square root of S^*S. Extend the definition of trace class to non-positive operators by saying that S is of trace class if $T = |S|$ is such that $tr(T)$ is finite. Show that the trace class operators form an ideal in $B(H, H)$.

16. Show that $T \in B(H, H)$ is a trace class operator if and only if $T = UV$ where U and V are Hilbert-Schmidt operators.

17. Derive a spectral theorem for compact normal operators.

18. Define the operator $T : L_2(0, 1) \to L_2(0, 1)$ by

$$Tu(x) = \int_0^x u(y)\, dy \; .$$

Show that T is compact, and find the eigenvalues of the self-adjoint compact operator T^*T. [Hint: T^* involves integration, so differentiate twice to get a second order ODE with two boundary conditions.]

19. For the differential operator

$$L = D^2 + xD \; ,$$

find a multiplying factor w so that wL is formally self adjoint. Find boundary conditions on $I = [0, 1]$ which make this operator into a regular Sturm-Liouville problem for which 0 is not an eigenvalue.

20. Give conditions under which the Sturm-Liouville operator

$$L = DpD + q \ ,$$

defined over an interval $I = [a, b]$, is a positive operator.

21. Write the Euler operator

$$L = x^2D^2 + xD$$

with the boundary conditions $u(1) = u(e) = 0$ on the interval $[1, e]$ as a regular Sturm-Liouville problem with an appropriate weight function w. Find the eigenvalues and eigenfunctions for this problem.

CHAPTER 4

Distributions

The theory of distributions, of "generalized functions," provides a general setting within which differentiation may be understood and exploited. It underlies the modern study of differential equations, optimization, the calculus of variations, and any subject utilizing differentiation.

4.1. The notion of generalized functions

The classic definition of the derivative is rather restrictive. For example, consider the function defined by

$$f(x) = \begin{cases} x \ , & x \geq 0 \ , \\ 0 \ , & x < 0 \ . \end{cases}$$

Then $f \in C^0(-\infty, \infty)$ and f is differentiable at every point except 0. The derivative of f is the Heaviside function

$$H(x) = \begin{cases} 1 \ , & x > 0 \ , \\ 0 \ , & x < 0 \ . \end{cases} \tag{4.1}$$

The nondifferentiability of f at 0 creates no particular problem, so should we consider f differentiable on $(-\infty, \infty)$? The derivative of H is also well defined, except at 0. However, it would appear that

$$H'(x) = \begin{cases} 0 \ , & x \neq 0 \ , \\ +\infty \ , & x = 0 \ , \end{cases}$$

at least in some sense. Can we make a precise statement? That is, can we generalize the notion of function so that H' is well defined?

We can make a precise statement if we use integration by parts. Recall that if $u, \phi \in C^1([a,b])$, then

$$\int_a^b u'\phi \, dx = u\phi\big|_a^b - \int_a^b u\phi' \, dx \ .$$

If $\phi \in C^1$ but $u \in C^0 \setminus C^1$, we can define "$\int_a^b u'v \, dx$" by the expression

$$u\phi\big|_a^b - \int_a^b u\phi' \, dx \ .$$

If we have enough "test functions" $\phi \in C^1$, then we can determine properties of u'. In practice, we take $\phi \in C_0^\infty(-\infty, \infty) = \{\psi \in C^\infty(-\infty, \infty) : \exists \ R > 0 \text{ such that } \psi(x) = 0 \ \forall \ |x| > R\}$ so that the boundary terms vanish for $a \to -\infty$, $b \to \infty$.

In our example, we have for all $\phi \in C_0^\infty$,

$$\begin{aligned}\int_{-\infty}^{\infty} f'\phi\,dx &\equiv -\int_{-\infty}^{\infty} f\phi'\,dx\\ &= -\int_0^{\infty} x\phi'\,dx\\ &= -x\phi\big|_0^\infty + \int_0^\infty \phi\,dx\\ &= \int_{-\infty}^{\infty} H\phi\,dx\ .\end{aligned}$$

Thus, we identify $f' = H$. Moreover,

$$\text{“}\int_{-\infty}^{\infty} H'\phi\,dx\text{”} \equiv -\int_{-\infty}^{\infty} H\phi'\,dx = -\int_0^\infty \phi'\,dx = \phi(0)\ ,$$

and we identify H' with evaluation at the origin! We call $H'(x) = \delta_0(x)$ the Dirac delta function. It is essentially zero everywhere except at the origin, where it must be infinite in some sense. It is *not* a function; it is a *generalized* function (or *distribution*).

We can continue. For example

$$\text{“}\int H''\phi\,dx\text{”} = \text{“}\int \delta_0'\phi\,dx\text{”} = \text{“}-\int \delta_0\phi'\,dx\text{”} = -\phi'(0)\ .$$

Obviously, $H'' = \delta_0'$ has no well defined value at the origin; nevertheless, we have a precise statement of the “integral” of δ_0' times any test function $\phi \in C_0^\infty$.

What we have described above can be viewed as a duality pairing between function spaces. That is, if we let

$$\mathcal{D} = C_0^\infty(-\infty,\infty)$$

be a space of test functions, then

$$f, \quad f' = H\ , \quad H' = \delta_0\ , \quad H'' = \delta_0'$$

can be viewed as linear functionals on $\mathcal{D}$, since integrals are linear and map to $\mathbb{F}$. For any linear functional u, we imagine

$$u(\phi) = \text{“}\int u\phi\,dx\ ,\text{”}$$

even when the integral is not defined in the Lebesgue sense, and define the derivative of u by

$$u'(\phi) = -u(\phi')\ .$$

Then also

$$u''(\phi) = -u'(\phi') = u(\phi'')\ ,$$

and so on for higher derivatives. In our case, precise statements are

$$\begin{aligned} f(\phi) &= \int f\phi\, dx \ , \\ f'(\phi) &= -f(\phi') = -\int f\phi'\, dx = \int H\phi\, dx = H(\phi) \ , \\ H'(\phi) &= -H(\phi') = -\int H\phi'\, dx = \phi(0) = \delta_0(\phi) \ , \\ H''(\phi) &= H(\phi'') = \int H\phi''\, dx = -\phi'(0) = -\delta_0(\phi') = \delta_0'(\phi) \ , \end{aligned}$$

for any $\phi \in \mathcal{D}$, repeating the integration by parts arguments for the integrals in the second line (which are now well defined).

We often wish to consider limit processes. To do so in this context would require that the linear functionals be continuous. That is, we require a topology on $\mathcal{D}$. Unfortunately, no simple topology will suffice.

4.2. Test Functions

Let $\Omega \subset \mathbb{R}^d$ be a domain, i.e., an open subset.

DEFINITION. If $f \in C^0(\Omega)$, the *support* of f is

$$\operatorname{supp}(f) = \overline{\{x \in \Omega : |f(x)| > 0\}} \subset \Omega \ ,$$

the *closure* (in Ω) of the set where f is nonzero. A *multi-index* $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{N}^d$ is an ordered d-tuple of nonnegative integers, and

$$|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d \ .$$

We let

$$\partial^\alpha = D^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d}$$

be a differential operator of order $|\alpha|$. Then we can define

$$\begin{aligned} C^n(\Omega) &= \{f \in C^0(\Omega) : D^\alpha f \in C^0(\Omega) \ \text{ for all } \ |\alpha| \le n\} \ , \\ C^\infty(\Omega) &= \{f \in C^0(\Omega) : D^\alpha f \in C^0(\Omega) \ \text{ for all } \ \alpha\} = \bigcap_{n=1}^{\infty} C^n(\Omega) \ , \\ \mathcal{D}(\Omega) &= C_0^\infty(\Omega) = \{f \in C^\infty(\Omega) : \operatorname{supp}(f) \text{ is compact}\}, \end{aligned}$$

and, if $K \subset\subset \Omega$ (i.e., K compact and $K \subset \Omega$),

$$\mathcal{D}_K = \{f \in C_0^\infty(\Omega) : \operatorname{supp}(f) \subset K\} \ .$$

PROPOSITION 4.1. *The sets $C^n(\Omega)$, $C^\infty(\Omega)$, $\mathcal{D}(\Omega)$, and $\mathcal{D}_K$ (for any $K \subset\subset \Omega$ with nonempty interior) are nonempty vector spaces.*

PROOF. It is trivial to verify that addition of functions and scalar multiplication are algebraically closed operations. Thus, each set is a vector space.

To see that these spaces are nonempty, we construct an element of $\mathcal{D}_K \subset \mathcal{D}(\Omega) \subset C^\infty(\Omega) \subset C^n(\Omega)$. Consider first *Cauchy's infinitely differentiable function* $\psi : \mathbb{R} \to \mathbb{R}$ given by

$$\psi(x) = \begin{cases} e^{-1/x^2} \ , & x > 0, \\ 0 \ , & x \le 0. \end{cases} \tag{4.2}$$

This function is clearly infinitely differentiable for $x \neq 0$, and its m^{th} derivative takes the form

$$\psi^{(m)}(x) = \begin{cases} R_m(x) e^{-1/x^2} \ , & x > 0, \\ 0 \ , & x < 0, \end{cases}$$

for some polynomial divided by x to a power $R_m(x)$. But L'Hôpital's rule implies that

$$\lim_{x \to 0} R_m(x) e^{-1/x^2} = 0 \ ,$$

so in fact $\psi^{(m)}$ is continuous at 0 for all m, and thus ψ is infinitely differentiable.

Now let $\phi(x) = \psi(1-x)\psi(1+x)$. Then $\phi \in C_0^\infty(\mathbb{R})$ and $\operatorname{supp}(\phi) = [-1,1]$. Finally, for $x \in \mathbb{R}^d$,

$$\Phi(x) = \phi(x_1)\phi(x_2)\dots\phi(x_d) \in C^\infty(\mathbb{R}^d)$$

has support $[-1,1]^d$. By translation and dilation, we can construct an element of $\mathcal{D}_K$. □

Corollary 4.2. *There exist nonanalytic functions.*

That is, there are functions not given by their Taylor series, since the Taylor series of $\psi(x)$ about 0 is 0, but $\psi(x) \neq 0$ for $x > 0$.

We define a norm on $C^n(\Omega)$ by

$$\|\phi\|_{n,\infty,\Omega} = \sum_{|\alpha| \le n} \|D^\alpha \phi\|_{L_\infty(\Omega)} \ .$$

Note that if $m \ge n$, then $\|\phi\|_{m,\infty,\Omega} \ge \|\phi\|_{n,\infty,\Omega}$, so we have a nested sequence of norms. We will use these to define convergence in $\mathcal{D}(\Omega)$, but we must be careful, as the following example shows.

Example. Take any $\phi \in C_0^\infty(\mathbb{R})$ such that $\operatorname{supp}(\phi) = [0,1]$ and $\phi(x) > 0$ for $x \in (0,1)$ (for example, we can construct such a function using Cauchy's infinitely differentiable function (4.2)). Define for any integer $n \ge 1$

$$\psi_n(x) = \sum_{j=1}^{n} \frac{1}{j} \phi(x-j) \in C_0^\infty(\mathbb{R}) \ ,$$

for which $\operatorname{supp}(\psi_n) = [1, n+1]$. Define also

$$\psi(x) = \sum_{j=1}^{\infty} \frac{1}{j} \phi(x-j) \in C^\infty(\mathbb{R}) \setminus C_0^\infty(\mathbb{R}) \ .$$

Now it is easy to verify that for any $m \ge 0$,

$$D^m \psi_n \xrightarrow{L_\infty} D^m \psi \ ;$$

that is,

$$\|\psi_n - \psi\|_{m,\infty,\mathbb{R}} \to 0$$

for each m, but $\psi \notin C_0^\infty(\mathbb{R})$.

To insure that $\mathcal{D}(\Omega)$ be complete, we will need both uniform convergence and a condition to force the limit to be compactly supported. The following definition suffices, and gives the usual topology on $C_0^\infty(\Omega)$, which we denote by $\mathcal{D} = \mathcal{D}(\Omega)$.

Definition. Let $\Omega \subset \mathbb{R}^d$ be a domain. We denote by $\mathcal{D}(\Omega)$ the vector space $C_0^\infty(\Omega)$ endowed with the following notion of convergence: A sequence $\{\phi_j\}_{j=1}^\infty \subset \mathcal{D}(\Omega)$ *converges* to $\phi \in \mathcal{D}(\Omega)$ if and only if there is some fixed $K \subset\subset \Omega$ such that $\operatorname{supp}(\phi_j) \subset K$ for all j and

$$\lim_{j\to\infty} \|\phi_j - \phi\|_{n,\infty,\Omega} = 0$$

for all n. Moreover, the sequence is *Cauchy* if $\operatorname{supp}(\phi_j) \subset K$ for all j for some fixed $K \subset\subset \Omega$ and, given $\epsilon > 0$ and $n \geq 0$, there exists $N > 0$ such that for all $j, k \geq N$,

$$\|\phi_j - \phi_k\|_{n,\infty,\Omega} \leq \epsilon.$$

That is, we have convergence if the ϕ_j are all localized to a compact set K, and each of their derivatives converges uniformly. Our definition does not identify open and closed sets; nevertheless, it does define a topology on $\mathcal{D}$. Unfortunately, $\mathcal{D}$ is *not* metrizable! However, it is easy to show and left to the reader that $\mathcal{D}(\Omega)$ is complete.

Theorem 4.3. *The linear space $\mathcal{D}(\Omega)$ is complete.*

4.3. Distributions

It turns out that, even though $\mathcal{D}(\Omega)$ is not a metric space, continuity and sequential continuity are equivalent for linear functionals. We do not use or prove the following fact, but it does explain our terminology.

Theorem 4.4. *If $T : \mathcal{D}(\Omega) \to \mathbb{F}$ is linear, then T is continuous if and only if T is sequentially continuous.*

Definition. A *distribution* or *generalized function* on a domain Ω is a (sequentially) continuous linear functional on $\mathcal{D}(\Omega)$. The vector space of all distributions is denoted $\mathcal{D}'(\Omega)$ (or $\mathcal{D}(\Omega)^*$). When $\Omega = \mathbb{R}^d$, we often write $\mathcal{D}$ for $\mathcal{D}(\mathbb{R}^d)$ and $\mathcal{D}'$ for $\mathcal{D}'(\mathbb{R}^d)$.

As in any linear space, we have the following result.

Theorem 4.5. *If $T : \mathcal{D}(\Omega) \to \mathbb{F}$ is linear, then T is sequentially continuous if and only if T is sequentially continuous at $0 \in \mathcal{D}$.*

We recast this result in our case as follows.

Theorem 4.6. *Suppose that $T : \mathcal{D}(\Omega) \to \mathbb{F}$ is linear. Then $T \in \mathcal{D}'(\Omega)$ (i.e., T is continuous) if and only if for every $K \subset\subset \Omega$, there are $n \geq 0$ and $C > 0$ such that*

$$|T(\phi)| \leq C\|\phi\|_{n,\infty,\Omega}$$

for every $\phi \in \mathcal{D}_K$.

Proof. Suppose that $T \in \mathcal{D}'(\Omega)$, but suppose also that the conclusion is false. Then there is some $K \subset\subset \Omega$ such that for every $n \geq 0$ and $m > 0$, we have some $\phi_{n,m} \in \mathcal{D}_K$ such that

$$|T(\phi_{n,m})| > m\|\phi_{n,m}\|_{n,\infty,\Omega} .$$

Normalize by setting $\hat{\phi}_{n,m} = \phi_{n,m}/(m\|\phi_{n,m}\|_{n,\infty,\Omega}) \in \mathcal{D}_K$. Then $|T(\hat{\phi}_{j,j})| > 1$, but $\hat{\phi}_{j,j} \to 0$ in $\mathcal{D}(\Omega)$ (since $\|\hat{\phi}_{j,j}\|_{n,\infty,\Omega} \leq \|\hat{\phi}_{j,j}\|_{j,\infty,\Omega} = 1/j$ for $j \geq n$), contradicting the hypothesis.

For the converse, suppose that $\phi_j \to 0$ in $\mathcal{D}(\Omega)$. Then there is some $K \subset\subset \Omega$ such that $\operatorname{supp}(\phi_j) \subset K$ for all j, and, by hypothesis, some n and C such that

$$|T(\phi_j)| \le C\|\phi_j\|_{n,\infty,\Omega} \to 0.$$

That is, T is (sequentially) continuous at 0. □

We proceed by giving some important examples.

DEFINITION.

$$L_{1,\text{loc}}(\Omega) = \left\{ f : \Omega \to \mathbb{F} \mid f \text{ is measurable and for every } K \subset\subset \Omega\ ,\ \int_K |f(x)|\,dx < \infty \right\}\ .$$

Note that $L_1(\Omega) \subset L_{1,\text{loc}}(\Omega)$. Any polynomial is in $L_{1,\text{loc}}(\Omega)$ but *not* in $L_1(\Omega)$, if Ω is unbounded. Elements of $L_{1,\text{loc}}(\Omega)$ may not be too singular at a point, but they may grow at infinity.

EXAMPLE. If $f \in L_{1,\text{loc}}(\Omega)$, we define $\Lambda_f \in \mathcal{D}'(\Omega)$ by

$$\Lambda_f(\phi) = \int_\Omega f(x)\phi(x)\,dx$$

for every $\phi \in \mathcal{D}(\Omega)$. Now Λ_f is obviously a linear functional; it is also continuous, since for $\phi \in \mathcal{D}_K$,

$$|\Lambda_f(\phi)| \le \int_K |f(x)|\,|\phi(x)|\,dx \le \left(\int_K |f(x)|\,dx \right) \|\phi\|_{0,\infty,\Omega}$$

satisfies the requirement of Theorem 4.6.

The mapping $f \mapsto \Lambda_f$ is one to one in the following sense.

PROPOSITION 4.7 (Lebesgue Lemma). *Let $f, g \in L_{1,\text{loc}}(\Omega)$. Then $\Lambda_f = \Lambda_g$ if and only if $f = g$ almost everywhere.*

PROOF. If $f = g$ a.e., then obviously $\Lambda_f = \Lambda_g$. Conversely, suppose $\Lambda_f = \Lambda_g$. Then $\Lambda_{f-g} = 0$ by linearity. Let

$$R = \{x \in \mathbb{R}^d : a_i \le x \le b_i,\ i = 1, \dots, d\} \subset \Omega$$

be an arbitrary closed rectangle, and let $\psi(x)$ be Cauchy's infinitely differentiable function on $\mathbb{R}$ given by (4.2). For $\varepsilon > 0$, let

$$\phi_\varepsilon(x) = \psi(\varepsilon - x)\psi(x) \ge 0$$

and

$$\Phi_\varepsilon(x) = \frac{\displaystyle\int_{-\infty}^x \phi_\varepsilon(\xi)\,d\xi}{\displaystyle\int_{-\infty}^\infty \phi_\varepsilon(\xi)\,d\xi}\ .$$

Then $\operatorname{supp}(\phi_\varepsilon) = [0, \varepsilon]$, $0 \le \Phi_\varepsilon(x) \le 1$, $\Phi_\varepsilon(x) = 0$ for $x \le 0$, and $\Phi_\varepsilon(x) = 1$ for $x \ge \varepsilon$.

Now let

$$\Psi_\varepsilon(x) = \prod_{i=1}^d \Phi_\varepsilon(x_i - a_i)\Phi_\varepsilon(b_i - x_i) \in \mathcal{D}_R\ .$$

If we let the characteristic function of R be

$$\chi_R(x) = \begin{cases} 1 \ , & x \in R \ , \\ 0 \ , & x \notin R \ , \end{cases}$$

then, pointwise, $\Psi_\varepsilon(x) \to \chi_R(x)$ as $\varepsilon \to 0$, and Lebesgue's Dominated Convergence Theorem implies that

$$(f-g)\Psi_\varepsilon \to (f-g)\chi_R$$

in $L_1(R)$. Thus

$$0 = \Lambda_{f-g}(\Psi_\varepsilon) = \int_R (f-g)(x)\Psi_\varepsilon(x)\,dx \to \int_R (f-g)(x)\,dx$$

as $\varepsilon \to 0$. So the integral of $f-g$ vanishes over any closed rectangle. From the theory of Lebesgue integration, we conclude that $f-g=0$, a.e. $\square$

We identify $f \in L_{1,\text{loc}}(\Omega)$ with $\Lambda_f \in \mathcal{D}'(\Omega)$, calling the *function* f a *distribution* in this sense. Since there are distributions that do not arise this way, as we will see, we call distributions *generalized functions*: functions are distributions but also more general objects are distributions.

DEFINITION. For $T \in \mathcal{D}'(\Omega)$, if there is $f \in L_{1,\text{loc}}(\Omega)$ such that $T = \Lambda_f$, then we call T a *regular* distribution. Otherwise T is a *singular* distribution.

Because the action of regular distributions is given by integration, people sometimes write, improperly but conveniently,

$$T(\phi) = \int_\Omega T\phi\,dx$$

for $T \in \mathcal{D}'(\Omega)$, $\phi \in \mathcal{D}(\Omega)$. To be more precise, we will often write

$$T(\phi) = \langle T, \phi \rangle = \langle T, \phi \rangle_{\mathcal{D}',\mathcal{D}} \ ,$$

where the notation $\langle \cdot, \cdot \rangle$ emphasizes the dual nature of the pairing of elements of $\mathcal{D}'(\Omega)$ and $\mathcal{D}(\Omega)$ and is sometimes, but not always, ordinary integration on Ω (i.e., the standard $L_2(\Omega)$ inner product).

EXAMPLE. We let $\delta_0 \in \mathcal{D}'(\Omega)$ be defined by

$$\langle \delta_0, \phi \rangle = \phi(0)$$

for every $\phi \in \mathcal{D}(\Omega)$. Again, linearity is trivial, and

$$|\langle \delta_0, \phi \rangle| = |\phi(0)| \le \|\phi\|_{0,\infty,\Omega}$$

implies by Theorem 4.6 that δ_0 is continuous. We call δ_0 the *Dirac mass, distribution* or *delta function* at 0. There is clearly no $f \in L_{1,\text{loc}}(\Omega)$ such that $\delta_0 = \Lambda_f$, so δ_0 is a singular distribution. If $x \in \Omega$, we also have $\delta_x \in \mathcal{D}'(\Omega)$ defined by

$$\langle \delta_x, \phi \rangle = \phi(x) \ .$$

This is the Dirac mass at x. This generalized function is often written, improperly, as

$$\delta_x(\xi) = \delta_0(\xi - x) = \delta_0(x - \xi) \ .$$

REMARK. We sketch a proof that $\mathcal{D}(\Omega)$ is *not* metrizable. The details are left to the reader. For $K \subset\subset \Omega$,

$$\mathcal{D}_K = \bigcap_{x\in\Omega\setminus K} \ker(\delta_x) \ .$$

Since $\ker(\delta_x)$ is closed, so is $\mathcal{D}_K$ (in $\mathcal{D}(\Omega)$). It is easy to show that $\mathcal{D}_K$ has empty interior in $\mathcal{D}$. But for a sequence $K_1 \subset K_2 \subset \cdots \subset \Omega$ of compact sets such that

$$\bigcup_{n=1}^{\infty} K_n = \Omega \ ,$$

we have

$$\mathcal{D}(\Omega) = \bigcup_{n=1}^{\infty} \mathcal{D}_{K_n} \ .$$

Apply the Baire Theorem to conclude that $\mathcal{D}(\Omega)$ is *not* metrizable.

EXAMPLE. If μ is either a complex Borel measure on Ω or a positive measure on Ω such that $\mu(K) < \infty$ for every $K \subset\subset \Omega$, then

$$\Lambda_\mu(\phi) = \int_\Omega \phi(x)\, d\mu(x)$$

defines a distribution, since

$$|\Lambda_\mu(\phi)| \le \mu(\operatorname{supp}(\phi))\|\phi\|_{0,\infty,\Omega} \ .$$

EXAMPLE. We define a distribution $PV\frac{1}{x} \in \mathcal{D}'(\mathbb{R})$ by

$$\langle PV\frac{1}{x}, \phi\rangle = PV\int \frac{1}{x}\phi(x)\, dx \equiv \lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \frac{1}{x}\phi(x)\, dx \ ,$$

called Cauchy's principle value of $1/x$. Since $1/x \notin L_{1,\mathrm{loc}}(\mathbb{R})$, we must verify that the limit is well defined. Fix $\phi \in \mathcal{D}$. Then integration by parts gives

$$\int_{|x|>\varepsilon} \frac{1}{x}\phi(x)\, dx = [\phi(-\varepsilon) - \phi(\varepsilon)]\ln\varepsilon - \int_{|x|>\varepsilon} \ln|x|\phi'(x)\, dx \ .$$

The boundary terms tend to 0:

$$\lim_{\varepsilon\downarrow 0}[\phi(-\varepsilon) - \phi(\varepsilon)]\ln\varepsilon = \lim_{\varepsilon\downarrow 0} 2\frac{\phi(-\varepsilon)-\phi(\varepsilon)}{2\varepsilon}\varepsilon\ln\varepsilon = -\phi'(0)\lim_{\varepsilon\downarrow 0}\varepsilon\ln\varepsilon = 0 \ .$$

Thus, if $\operatorname{supp}(\phi) \subset [-R, R] = K$, then

$$PV\int \frac{1}{x}\phi(x) = -\lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \ln|x|\phi'(x)\, dx = -\int_{-R}^{R} \ln|x|\phi'(x)\, dx$$

exists, and

$$\Big|PV\int \frac{1}{x}\phi(x)\, dx\Big| \le \left(\int_{-R}^{R} \big|\ln|x|\big|\, dx\right)\|\phi\|_{1,\infty,\mathbb{R}}$$

shows that $PV(1/x)$ is a distribution, since the latter integral is finite.

4.4. Operations with distributions

A simple way to define a new distribution from an existing one is to use duality. If $T : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is sequentially continuous and linear, then $T^* : \mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ satisfies

$$\langle u, T\phi \rangle = \langle T^* u, \phi \rangle$$

for all $u \in \mathcal{D}'(\Omega)$, $\phi \in \mathcal{D}(\Omega)$. Obviously $T^* u = u \circ T$ is sequentially continuous and linear.

PROPOSITION 4.8. *If $u \in \mathcal{D}'(\Omega)$ and $T : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is sequentially continuous and linear, then $T^* u = u \circ T \in \mathcal{D}'(\Omega)$.*

We use this proposition below to conclude that our linear functionals are distributions; alternatively, we could have shown the condition of Theorem 4.6, as the reader can verify.

4.4.1. Multiplication by a smooth function. If $f \in C^\infty(\Omega)$, we can define $T_f : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ by $T_f(\phi) = f\phi$. Obviously T_f is linear and sequentially continuous, by the product rule for differentiation. Thus, for any $u \in \mathcal{D}'(\Omega)$, $T_f^* u = u \circ T_f \in \mathcal{D}'(\Omega)$. But if $u = \Lambda_u$ is a regular distribution (i.e., $u \in L_{1,\mathrm{loc}}(\Omega)$),

$$\begin{aligned} \langle T_f^* u, \phi \rangle &= \langle u, T_f \phi \rangle = \langle u, f\phi \rangle \\ &= \int_\Omega u(x) f(x) \phi(x)\, dx \\ &= \langle fu, \phi \rangle \ , \end{aligned}$$

for any $\phi \in \mathcal{D}(\Omega)$. We define for any $u \in \mathcal{D}'$ and $f \in C^\infty(\Omega)$ a new distribution, denoted fu, as $fu = T_f^* u$, satisfying

$$\langle fu, \phi \rangle = \langle u, f\phi \rangle \qquad \forall\ \phi \in \mathcal{D}(\Omega) \ .$$

Thus we can multiply any distribution by a smooth function, and

$$f\Lambda_u = \Lambda_{fu}$$

for a regular distribution.

4.4.2. Differentiation. Our most important example is differentiation. Note that $D^\alpha : \mathcal{D}(\Omega) \to \mathcal{D}(\Omega)$ is sequentially continuous for any multi-index α, so $(D^\alpha)^* u = u \circ D^\alpha \in \mathcal{D}'(\Omega)$. Moreover, for $\phi, \psi \in C_0^\infty(\Omega)$,

$$\int D^\alpha \phi(x) \psi(x)\, dx = (-1)^{|\alpha|} \int \phi(x) D^\alpha \psi(x)\, dx \ ,$$

using integration by parts.

DEFINITION. If α is a multi-index and $u \in \mathcal{D}'(\Omega)$, we define $D^\alpha u \in \mathcal{D}'(\Omega)$ by

$$\langle D^\alpha u, \phi \rangle = (-1)^{|\alpha|} \langle u, D^\alpha \phi \rangle \qquad \forall\ \phi \in \mathcal{D}(\Omega) \ . \tag{4.3}$$

We should verify that this definition is consistent with our usual notion of differentiation when $u = \Lambda_u$ is a regular distribution.

PROPOSITION 4.9. *Suppose $u \in C^n(\Omega)$ for $n \geq 0$. Let α be a multi-index such that $|\alpha| \leq n$, and denote the classical α-partial derivatives of u by $\partial^\alpha u = \partial^\alpha u / \partial x^\alpha$. Then*

$$D^\alpha u \equiv D^\alpha \Lambda_u = \partial^\alpha u \ .$$

That is, the two distributions $D^\alpha \Lambda_u$ and $\Lambda_{\partial^\alpha u}$ agree.

Proof. For any $\phi \in \mathcal{D}(\Omega)$,

$$\begin{aligned}\langle D^\alpha \Lambda_u, \phi\rangle &= (-1)^{|\alpha|}\langle \Lambda_u, D^\alpha\phi\rangle \\ &= (-1)^{|\alpha|}\int u(x) D^\alpha \phi(x)\,dx \\ &= \int \partial^\alpha u(x)\phi(x)\,dx \\ &= \langle \partial^\alpha u, \phi\rangle\ ,\end{aligned}$$

where the third equality comes by the ordinary integration by parts formula. Since ϕ is arbitrary, $D^\alpha\Lambda_u = \partial^\alpha u$. □

Example. If $H(x)$ is the Heaviside function (4.1), then $H \in L_{1,\mathrm{loc}}(\mathbb{R})$ is also a distribution, and, for any $\phi \in \mathcal{D}(\mathbb{R})$,

$$\begin{aligned}\langle H', \phi\rangle &= -\langle H, \phi'\rangle \\ &= -\int_{-\infty}^{\infty} H(x)\phi'(x)\,dx \\ &= -\int_0^\infty \phi'(x)\,dx \\ &= \phi(0) = \langle \delta_0, \phi\rangle\ .\end{aligned}$$

Thus $H' = \delta_0$, as *distributions*.

Example. Since $\ln|x| \in L_{1,\mathrm{loc}}(\mathbb{R})$ is a distribution, the distributional derivative applied to $\phi \in \mathcal{D}$ is

$$\begin{aligned}\langle D\ln|x|, \phi\rangle &= -\langle \ln|x|, D\phi\rangle \\ &= -\int \ln|x|\phi'(x)\,dx \\ &= -\lim_{\varepsilon\downarrow 0}\int_{|x|>\varepsilon} \ln|x|\phi'(x)\,dx \\ &= \lim_{\varepsilon\downarrow 0}\left\{\int_{|x|>0}\frac{1}{x}\phi(x)\,dx + (\phi(\varepsilon)-\phi(-\varepsilon))\ln|\varepsilon|\right\} \\ &= \lim_{\varepsilon\downarrow 0}\int_{|x|>0}\frac{1}{x}\phi(x)\,dx\ .\end{aligned}$$

Thus $D\ln|x| = PV(1/x)$.

Proposition 4.10. *If $u \in \mathcal{D}'(\Omega)$ and α and β are multi-indices, then*

$$D^\alpha D^\beta u = D^\beta D^\alpha u = D^{\alpha+\beta} u\ .$$

Proof. For $\phi \in \mathcal{D}(\Omega)$,

$$\begin{aligned}\langle D^\alpha D^\beta u, \phi\rangle &= (-1)^{|\alpha|}\langle D^\beta u, D^\alpha\phi\rangle \\ &= (-1)^{|\alpha|+|\beta|}\langle u, D^\beta D^\alpha \phi\rangle \\ &= (-1)^{|\beta|+|\alpha|}\langle u, D^\alpha D^\beta\phi\rangle\ .\end{aligned}$$

Thus α and β may be interchanged. Moreover,

$$\begin{aligned}\langle D^\alpha D^\beta u, \phi\rangle &= (-1)^{|\alpha|+|\beta|}\langle u, D^\alpha D^\beta \phi\rangle \\ &= (-1)^{|\alpha+\beta|}\langle u, D^{\alpha+\beta}\phi\rangle \\ &= \langle D^{\alpha+\beta} u, \phi\rangle \ . \qquad \square\end{aligned}$$

LEMMA 4.11 (Leibniz Rule). *Let $f \in C^\infty(\Omega)$, $u \in \mathcal{D}'(\Omega)$, and α a multi-index. Then*

$$D^\alpha(fu) = \sum_{\beta\le\alpha} \binom{\alpha}{\beta} D^{\alpha-\beta} f D^\beta u \in \mathcal{D}' \ ,$$

where

$$\binom{\alpha}{\beta} = \frac{\alpha!}{(\alpha-\beta)!\beta!} \ ,$$

$\alpha! = \alpha_1!\alpha_2!\cdots\alpha_d!$, and $\beta \le \alpha$ means that β is a multi-index with $\beta_i \le \alpha_i$ for $i = 1, \dots, d$.

If $u \in C^\infty(\Omega)$, this is just the product rule for differentiation.

PROOF. By the previous proposition, we have the theorem if it is true for multi-indices that have a single nonzero component, say the first component. We proceed by induction on $n = |\alpha|$. The result holds for $n = 0$, but we will need the result for $n = 1$. Denote D^α by D_1^n. When $n = 1$, for any $\phi \in \mathcal{D}(\Omega)$,

$$\begin{aligned}\langle D_1(fu), \phi\rangle &= -\langle fu, D_1\phi\rangle \\ &= -\langle u, fD_1\phi\rangle = -\langle u, D_1(f\phi) - D_1 f\phi\rangle \\ &= \langle D_1 u, f\phi\rangle + \langle u, D_1 f\phi\rangle \\ &= \langle fD_1 u + D_1 f u, \phi\rangle \ ,\end{aligned}$$

and the result holds.

Now assume the result for derivatives up to order $n-1$. Then

$$\begin{aligned}D_1^n(fu) &= D_1 D_1^{n-1}(fu) \\ &= D_1 \sum_{j=0}^{n-1} \binom{n-1}{j} D_1^{n-1-j} f D_1^j u \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} (D_1^{n-j} f D_1^j u + D_1^{n-1-j} f D_1^{j+1} u) \\ &= \sum_{j=0}^{n-1} \binom{n-1}{j} D_1^{n-j} f D_1^j u + \sum_{j=1}^{n} \binom{n-1}{j-1} D_1^{n-j} f D_1^j u \\ &= \sum_{j=0}^{n} \binom{n}{j} D_1^{n-j} f D_1^j u \ ,\end{aligned}$$

where the last equality follows from the combinatorial identity

$$\binom{n}{j} = \binom{n-1}{j} + \binom{n-1}{j-1} \ ,$$

and so the induction proceeds. $\square$

Example. Consider $f(x) = x \ln|x|$. Since $x \in C^\infty(\mathbb{R})$ and $\ln|x| \in \mathcal{D}'$, we have

$$D(x \ln|x|) = \ln|x| + x\, PV\left(\frac{1}{x}\right) .$$

But, for $\phi \in \mathcal{D}$, integration by parts gives

$$\begin{aligned}\langle D(x \ln|x|), \phi\rangle &= -\langle x \ln|x|, D\phi\rangle \\ &= -\int x \ln|x| \phi'(x)\, dx \\ &= \int_0^\infty (\ln|x| + 1)\phi(x)\, dx + \int_{-\infty}^0 (\ln|x| + 1)\phi(x)\, dx \\ &= \langle \ln|x| + 1, \phi\rangle \ .\end{aligned}$$

Thus

$$x\, PV\left(\frac{1}{x}\right) = 1 \ ,$$

which the reader can prove directly quite easily.

4.4.3. Translations and dilations of $\mathbb{R}^d$. Assume $\Omega = \mathbb{R}^d$ and define for any fixed $x \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$, $\lambda \neq 0$, the maps $\tau_x : \mathcal{D} \to \mathcal{D}$ and $T_\lambda : \mathcal{D} \to \mathcal{D}$ by

$$\tau_x \phi(y) = \phi(y - x) \quad \text{and} \quad T_\lambda \phi(y) = \phi(\lambda y) \ ,$$

for any $y \in \mathbb{R}^d$. These maps translate and dilate the domain. They are clearly sequentially continuous and linear maps on $\mathcal{D}$.

Given $u \in \mathcal{D}'$, we define the distributions $\tau_x u$ and $T_\lambda u$ for $\phi \in \mathcal{D}$ by

$$\begin{aligned}\langle \tau_x u, \phi\rangle &= \langle u, \tau_{-x}\phi\rangle \ , \\ \langle T_\lambda u, \phi\rangle &= \frac{1}{|\lambda|^d}\langle u, T_{1/\lambda}\phi\rangle \ .\end{aligned}$$

These definitions are clearly consistent with the usual change of variables formulas for integrals when u is a regular distribution.

4.4.4. Convolutions. If $f, g : \mathbb{R}^d \to \mathbb{F}$ are functions, we define the *convolution* of f and g, a function denoted $f * g : \mathbb{R}^d \to \mathbb{F}$, by

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

provided the (Lebesgue) integral exists for almost every $x \in \mathbb{R}^d$. If we let τ_x denote spatial translation and R denote reflection (i.e., $R = T_{-1}$ from the previous subsection), then

$$f * g(x) = \int_{\mathbb{R}^d} f(y)(\tau_x R g)(y)\, dy \ .$$

This motivates the definition of the *convolution* of a distribution $u \in \mathcal{D}'(\mathbb{R}^d)$ and a test function $\phi \in \mathcal{D}(\mathbb{R}^d)$:

$$(u * \phi)(x) = \langle u, \tau_x R\phi\rangle = \langle R\tau_{-x} u, \phi\rangle \ , \quad \text{for any} \quad x \in \mathbb{R}^d \ .$$

Indeed, $R\tau_{-x} u = u \circ \tau_x \circ R \in \mathcal{D}'$ is well defined.

Example. If $\phi \in \mathcal{D}$ and $x \in \mathbb{R}^d$, then

$$\delta_0 * \phi(x) = \langle \delta_0, \tau_x R\phi \rangle = \phi(x) \ .$$

If $u \in \mathcal{D}'$, then

$$u * R\phi(0) = \langle u, \phi \rangle \ .$$

Proposition 4.12. *If $u \in \mathcal{D}'(\mathbb{R}^d)$ and $\phi \in \mathcal{D}(\mathbb{R}^d)$, then*

(a) *for any $x \in \mathbb{R}^d$,*

$$\tau_x(u * \phi) = (\tau_x u) * \phi = u * (\tau_x \phi) \ ,$$

(b) *$u * \phi \in C^\infty(\mathbb{R}^d)$ and, for any multi-index α,*

$$D^\alpha(u * \phi) = (D^\alpha u) * \phi = u * (D^\alpha \phi) \ .$$

Remark. Since u could be a function in $L_{1,\mathrm{loc}}(\mathbb{R}^d)$, these results hold for functions as well.

Proof. For (a), note that

$$\begin{aligned}
\tau_x(u * \phi)(y) &= (u * \phi)(y - x) = \langle u, \tau_{y-x} R\phi \rangle \ , \\
(\tau_x u) * \phi(y) &= \langle \tau_x u, \tau_y R\phi \rangle = \langle u, \tau_{y-x} R\phi \rangle \ , \\
(u * \tau_x \phi)(y) &= \langle u, \tau_y R \tau_x \phi \rangle = \langle u, \tau_{y-x} R\phi \rangle \ .
\end{aligned}$$

Part of (b) is easy:

$$\begin{aligned}
D^\alpha u * \phi(x) &= \langle D^\alpha u, \tau_x R\phi \rangle \\
&= (-1)^{|\alpha|} \langle u, D^\alpha \tau_x R\phi \rangle \\
&= (-1)^{|\alpha|} \langle u, \tau_x D^\alpha R\phi \rangle \\
&= \langle u, \tau_x R D^\alpha \phi \rangle \\
&= u * D^\alpha \phi(x) \ .
\end{aligned}$$

Now for $h > 0$ and $\mathbf{e} \in \mathbb{R}^d$ a unit vector, let

$$T_h = \frac{1}{h}(I - \tau_{h\mathbf{e}}) \ .$$

Then

$$\lim_{h \to 0} T_h \phi(x) = \frac{\partial \phi}{\partial e}(x)$$

pointwise; in fact the convergence is uniform since $\partial\phi/\partial e$ is uniformly continuous (it has a bounded gradient). Given $\varepsilon > 0$, there is $\delta > 0$ such that

$$\Big| \frac{\partial \phi}{\partial e}(x) - \frac{\partial \phi}{\partial e}(y) \Big| \le \varepsilon$$

whenever $|x - y| < \delta$. Thus

$$\Big| T_h \phi(x) - \frac{\partial \phi}{\partial e}(x) \Big| = \Big| \frac{1}{h} \int_{-h}^0 \Big(\frac{\partial \phi}{\partial e}(x + s\mathbf{e}) - \frac{\partial \phi}{\partial e}(x) \Big)\, ds \Big| \le \varepsilon$$

whenever $|h| < \delta$. Similarly

$$D^\alpha T_h \phi = T_h D^\alpha \phi \xrightarrow{L_\infty(\mathbb{R}^d)} D^\alpha \frac{\partial \phi}{\partial e} \ ,$$

so we conclude that

$$T_h \phi \xrightarrow{\mathcal{D}} \frac{\partial \phi}{\partial e} \quad \text{as} \quad h \to 0 \ .$$

Now, by part (a), for any $x \in \mathbb{R}^d$,

$$T_h(u * \phi)(x) = u * T_h\phi(x) \ ,$$

so

$$\lim_{h\to 0} T_h(u * \phi)(x) = \lim_{h\to 0} u * T_h\phi(x) = u * \frac{\partial \phi}{\partial e}(x) \ ,$$

since $u \circ \tau_x \circ R \in \mathcal{D}'$. Thus $\frac{\partial}{\partial e}(u * \phi)$ exists and equals $u * \frac{\partial \phi}{\partial e}$. By iteration, (b) follows. $\square$

If $\phi, \psi \in \mathcal{D}$, then $\phi * \psi \in \mathcal{D}$, since

$$\operatorname{supp}(\phi * \psi) \subset \operatorname{supp}(\phi) + \operatorname{supp}(\psi) \ .$$

Proposition 4.13. *If $\phi, \psi \in \mathcal{D}$, $u \in \mathcal{D}'$, then*

$$(u * \phi) * \psi = u * (\phi * \psi) \ .$$

Proof. Since $\phi * \psi$ is uniformly continuous, we may approximate the convolution integral by a Riemann sum: for $h > 0$,

$$r_h(x) = \sum_{k\in\mathbb{Z}^d} \phi(x - kh)\psi(kh)h^d \ ,$$

and $r_h(x) \to \phi * \psi(x)$ uniformly in x as $h \to 0$. Moreover,

$$D^\alpha r_h \to (D^\alpha \phi) * \psi = D^\alpha(\phi * \psi)$$

uniformly, and

$$\operatorname{supp}(r_h) \subset \operatorname{supp}(\phi) + \operatorname{supp}(\psi) \ .$$

We conclude that

$$r_h \xrightarrow{\mathcal{D}} \phi * \psi \ .$$

Thus

$$\begin{aligned} u * (\phi * \psi)(x) &= \lim_{h\downarrow 0} u * r_h(x) \\ &= \lim_{h\downarrow 0} \sum_{k\in\mathbb{Z}^d} u * \phi(x - kh)\psi(kh)h^d \\ &= (u * \phi) * \psi(x) \ . \qquad \square \end{aligned}$$

4.5. Convergence of distributions and approximations to the identity

We endow $\mathcal{D}'(\Omega)$ with its weak topology. Although we will not prove or use the fact, $\mathcal{D}$ is reflexive, so the weak topology on $\mathcal{D}'(\Omega)$ is the weak-$*$ topology. The weak topology on $\mathcal{D}'(\Omega)$ is defined by the following notion of convergence: a sequence $\{u_j\}_{j=1}^\infty \subset \mathcal{D}'(\Omega)$ converges to $u \in \mathcal{D}'(\Omega)$ if and only if

$$\langle u_j, \phi\rangle \to \langle u, \phi\rangle \quad \forall\ \phi \in \mathcal{D}(\Omega) \ .$$

As the following proposition states, $\mathcal{D}'(\Omega)$ is (sequentially) complete.

Proposition 4.14. *If $\{u_n\}_{n=1}^\infty \subset \mathcal{D}'(\Omega)$ and $\{\langle u_n, \phi\rangle\}_{n=1}^\infty \subset \mathbb{F}$ is Cauchy for all $\phi \in \mathcal{D}(\Omega)$, then $u : \mathcal{D} \to \mathbb{F}$ defined by*

$$u(\phi) = \langle u, \phi\rangle = \lim_{n\to\infty} \langle u_n, \phi\rangle$$

defines a distribution.

The existence and linearity of u is clear. We hypothesize pointwise convergence, so the continuity of u follows from a uniform boundedness principle, which we do not prove here.

PROPOSITION 4.15. *If* $u_n \xrightarrow{\mathcal{D}'(\Omega)} u$ *and* α *is any multi-index, then* $D^\alpha u_n \xrightarrow{\mathcal{D}'(\Omega)} D^\alpha u$.

PROOF. For any $\phi \in \mathcal{D}$,

$$\langle D^\alpha u_n, \phi\rangle = (-1)^{|\alpha|}\langle u_n, D^\alpha \phi\rangle \to (-1)^{|\alpha|}\langle u, D^\alpha \phi\rangle = \langle D^\alpha u, \phi\rangle \ . \qquad \square$$

We leave the following two propositions as exercises.

PROPOSITION 4.16. *If* $u \in \mathcal{D}'(\Omega)$ *and* α *is a multi-index with* $|\alpha| = 1$, *then*

$$\lim_{h\to 0} \frac{1}{h}(\tau_{h\alpha} u - u) \xrightarrow{\mathcal{D}'(\Omega)} D^\alpha u \ ,$$

wherein the first α *is interpreted as a unit vector in* $\mathbb{R}^d$.

PROPOSITION 4.17. *Let* $\chi_R(x)$ *denote the characteristic function of* $R \subset \mathbb{R}$. *For* $\varepsilon > 0$,

$$\frac{1}{\varepsilon}\chi_{[-\varepsilon/2,\varepsilon/2]} \xrightarrow{\mathcal{D}'(\mathbb{R})} \delta_0$$

as $\varepsilon \to 0$.

DEFINITION. Let $\varphi \in \mathcal{D}(\mathbb{R}^d)$ satisfy
(a) $\varphi \geq 0$,
(b) $\int \varphi(x)\,dx = 1$,
and define for $\varepsilon > 0$

$$\varphi_\varepsilon(x) = \frac{1}{\varepsilon^d}\varphi\Big(\frac{x}{\varepsilon}\Big) \ .$$

Then we call $\{\varphi_\varepsilon\}_{\varepsilon>0}$ an *approximation to the identity*.

The following is easily verified.

PROPOSITION 4.18. *If* $\{\varphi_\varepsilon\}_{\varepsilon>0}$ *is an approximation to the identity, then*

$$\int \varphi_\varepsilon(x)\,dx = 1 \quad \forall\ \varepsilon > 0$$

and $\operatorname{supp}(\varphi_\varepsilon) \to \{0\}$ *as* $\varepsilon \to 0$.

THEOREM 4.19. *Let* $\{\varphi_\varepsilon\}_{\varepsilon>0}$ *be an approximation to the identity.*
(a) *If* $\psi \in \mathcal{D}$, *then* $\psi * \varphi_\varepsilon \xrightarrow{\mathcal{D}} \psi$.
(b) *If* $u \in \mathcal{D}'$, *then* $u * \varphi_\varepsilon \xrightarrow{\mathcal{D}'} u$.

Since $u * \varphi_\varepsilon \in C^\infty$, we see that $C^\infty(\mathbb{R}^d) \subset \mathcal{D}'$ is dense. Moreover, $\{\varphi_\varepsilon\}_{\varepsilon>0}$ approximates the convolution identity δ_0.

PROOF. (a) Let $\operatorname{supp}(\varphi) \subset \overline{B_R(0)}$ for some $R > 0$. First note that for $0 < \varepsilon \leq 1$,

$$\operatorname{supp}(\psi * \varphi_\varepsilon) \subset \operatorname{supp}(\psi) + \operatorname{supp}(\varphi_\varepsilon) \subset \operatorname{supp}(\psi) + \overline{B_R(0)} = K$$

is contained in a compact set. If $f \in C_0^\infty(\mathbb{R}^d)$, then

$$\begin{aligned} f * \varphi_\varepsilon(x) &= \int f(x-y)\varphi_\varepsilon(y)\,dy \\ &= \int f(x-y)\varepsilon^{-d}\varphi(\varepsilon^{-1}y)\,dy \\ &= \int f(x-\varepsilon z)\varphi(z)\,dz \\ &= \int (f(x-\varepsilon z) - f(x))\varphi(z)\,dz + f(x)\ , \end{aligned}$$

and this converges uniformly to $f(x)$. Thus for any multi-index α,

$$D^\alpha(\psi * \varphi_\varepsilon) = (D^\alpha\psi) * \varphi_\varepsilon \xrightarrow{L_\infty} D^\alpha\psi\ ;$$

that is, $\psi * \varphi_\varepsilon \xrightarrow{\mathcal{D}_K} \psi$, and so also $\psi * \varphi_\varepsilon \xrightarrow{\mathcal{D}} \psi$.

(b) Since convolution generates a (continuous) distribution for any fixed x, by (a) and Proposition 4.13, we have for $\psi \in \mathcal{D}$,

$$\begin{aligned} \langle u, \psi\rangle &= u * R\psi(0) \\ &= \lim_{\varepsilon\to 0} u * (R\psi * \varphi_\varepsilon)(0) \\ &= \lim_{\varepsilon\to 0} (u * \varphi_\varepsilon) * R\psi(0) \\ &= \lim_{\varepsilon\to 0} \langle u * \varphi_\varepsilon, \psi\rangle\ . \end{aligned}$$

□

Corollary 4.20. $\varphi_\varepsilon = \delta_0 * \varphi_\varepsilon \xrightarrow{\mathcal{D}'} \delta_0$.

4.6. Some Applications to Linear Differential Equations

An operator $L : C^m(\mathbb{R}^d) \to C^0(\mathbb{R}^d)$ is called a *linear differential operator* if there are functions $a_\alpha \in C^0(\mathbb{R}^d)$ for all multi-indices α such that

$$L = \sum_{|\alpha|\le m} a_\alpha D^\alpha\ . \tag{4.4}$$

The maximal $|\alpha|$ for which a_α is not identically zero is the *order* of L.

If $a_\alpha \in C^\infty(\mathbb{R}^d)$, then we can extend L to

$$L : \mathcal{D}' \to \mathcal{D}'\ ,$$

and this operator is linear and continuous. Given $f \in \mathcal{D}'$, we have the partial or ordinary differential equation

$$Lu = f \quad \text{in} \quad \mathcal{D}'$$

for which we seek a *distributional solution* $u \in \mathcal{D}'$ such that

$$\langle Lu, \phi\rangle = \langle f, \phi\rangle \quad \forall\ \phi \in \mathcal{D}\ .$$

We say that any such u is a *classical solution* if $u \in C^m(\mathbb{R}^d)$ satisfies the equation pointwise. If u is a regular distribution, then u is called a *weak solution* (so classical solutions are also weak solutions). Note that if $u \in \mathcal{D}'$ solves the equation, it would fail to be a weak solution if u is a singular distribution.

4.6.1. Ordinary differential equations. We consider the case when $d = 1$.

LEMMA 4.21. *Let $\phi \in \mathcal{D}(\mathbb{R})$. Then $\int \phi(x)\,dx = 0$ if and only if there is some $\psi \in \mathcal{D}(\mathbb{R})$ such that $\phi = \psi'$.*

The proof is left to the reader.

DEFINITION. A distribution $v \in \mathcal{D}'(\mathbb{R})$ is a *primitive* of $u \in \mathcal{D}'(\mathbb{R})$ if $Dv \equiv v' = u$.

THEOREM 4.22. *Every $u \in \mathcal{D}'(\mathbb{R})$ has infinitely many primitives, and any two differ by a constant.*

PROOF. Let

$$\begin{aligned} \mathcal{D}_0 &= \left\{\phi \in \mathcal{D}(\mathbb{R}) : \int \phi(x)\,dx = 0\right\} \\ &= \{\phi \in \mathcal{D}(\mathbb{R}) : \text{ there is } \psi \in \mathcal{D}(\mathbb{R}) \text{ such that } \psi' = \phi\} \ . \end{aligned}$$

Then $\mathcal{D}_0$ is a vector space and $v \in \mathcal{D}'$ is a primitive for u if and only if

$$\langle u, \psi \rangle = \langle v', \psi \rangle = -\langle v, \psi' \rangle \quad \forall\ \psi \in \mathcal{D} \ ;$$

that is, by the lemma, if and only if

$$\langle v, \phi \rangle = -\left\langle u, \int_{-\infty}^{x} \phi(\xi)\,d\xi \right\rangle \quad \forall\ \phi \in \mathcal{D}_0 \ .$$

Thus $v : \mathcal{D}_0 \to \mathbb{F}$ is defined. We extend v to $\mathcal{D}$ as follows. Fix $\phi_1 \in \mathcal{D}$ such that $\int \phi_1(x)\,dx = 1$. Then any $\psi \in \mathcal{D}$ is uniquely decomposed as

$$\psi = \phi + \langle 1, \psi \rangle \phi_1$$

where $\phi \in \mathcal{D}_0$. Choose $c \in \mathbb{F}$ and define v_c for $\psi \in \mathcal{D}$ by

$$\langle v_c, \psi \rangle = \langle v_c, \phi \rangle + \langle 1, \psi \rangle \langle v_c, \phi_1 \rangle \equiv \langle v, \phi \rangle + c\langle 1, \psi \rangle \ .$$

Clearly v_c is linear and $v_c|_{\mathcal{D}_0} = v$. We claim that v_c is continuous. If $\psi_n \xrightarrow{\mathcal{D}} 0$, then $\langle 1, \psi_n \rangle \xrightarrow{\mathbb{F}} 0$ and $\mathcal{D}_0 \ni \phi_n = \psi_n - \langle 1, \psi_n \rangle \phi_1 \xrightarrow{\mathcal{D}} 0$, as does $\int_{-\infty}^{x} \phi_n(\xi)\,d\xi$. Therefore $\langle v, \phi_n \rangle = -\langle u, \int_{-\infty}^{x} \phi_n(\xi)\,d\xi \rangle \to 0$, and so also $\langle v_c, \psi_n \rangle \to 0$. Thus v_c, for each $c \in \mathbb{F}$, is a distribution and $v_c' = u$.

If $v, w \in \mathcal{D}'$ are primitives of u, then for $\psi \in \mathcal{D}$ expanded as above with $\phi \in \mathcal{D}_0$,

$$\begin{aligned} \langle v - w, \psi \rangle &= \langle v - w, \phi \rangle + \langle v - w, \langle 1, \psi \rangle \phi_1 \rangle \\ &= 0 + \big\langle \langle v - w, \phi_1 \rangle, \psi \big\rangle \ , \end{aligned}$$

and so

$$v - w = \langle v - w, \phi_1 \rangle \in \mathbb{F} \ . \qquad \square$$

COROLLARY 4.23. *If $u' = 0$ in $\mathcal{D}'(\mathbb{R})$, then u is constant.*

COROLLARY 4.24. *If $a \in \mathbb{F}$, then $u' = au$ in $\mathcal{D}'(\mathbb{R})$ has only classical solutions given by*

$$u(x) = Ce^{ax}$$

for some $C \in \mathbb{F}$.

PROOF. We have the existence of at least the solutions Ce^{ax}. Let u be any distributional solution. Note that $e^{-ax} \in C^\infty(\mathbb{R})$, so $v = e^{-ax}u \in \mathcal{D}'$ and Leibniz rule implies

$$v' = -ae^{-ax}u + e^{-ax}u' = e^{-ax}(u' - au) = 0 \ .$$

Thus $v = C$, a constant, and $u = Ce^{ax}$. □

COROLLARY 4.25. *Let $a(x), b(x) \in C^\infty(\mathbb{R})$. Then the differential equation*

$$u' + a(x)u = b(x) \quad \text{in} \quad \mathcal{D}'(\mathbb{R}) \tag{4.5}$$

possesses only the classical solutions

$$u = e^{-A(x)}\left[\int_0^x e^{A(\xi)}b(\xi)\,d\xi + C\right]$$

for any $C \in \mathbb{F}$ where A is any primitive of a (i.e., $A' = a$).

PROOF. If $u, v \in \mathcal{D}'$ solve the equation, then their difference solves the homogeneous equation

$$w' + a(x)w = 0 \quad \text{in} \quad \mathcal{D}'(\mathbb{R}) \ .$$

But, similar to the proof above, such solutions have the form

$$w = Ce^{-A(x)}$$

(i.e., $(e^{A(x)}w)' = e^{A(x)}w' + a(x)e^{A(x)}w = 0$). Thus any solution of the nonhomogeneous equation (4.5) has the form

$$u = Ce^{-A(x)} + v$$

where v is any solution. Since

$$v = e^{-A(x)}\int_0^x e^{A(\xi)}b(\xi)\,d\xi$$

is a solution, the result follows. □

Not all equations are so simple.

EXAMPLE. Let us solve

$$xu' = 1 \quad \text{in} \quad \mathcal{D}'(\mathbb{R}) \ .$$

We know $u = \ln|x| \in L_{1,\text{loc}}(\mathbb{R})$ is a solution, since $(\ln|x|)' = PV(1/x)$ and $xPV(1/x) = 1$. All other solutions are given by adding any solution to

$$xv' = 0 \quad \text{in} \quad \mathcal{D}'(\mathbb{R}) \ .$$

Since $v' \in \mathcal{D}'(\mathbb{R})$ may *not* be a regular distribution, we must *not* divide by x to conclude v is a constant (since $x = 0$ is possible). In fact,

$$v = c_1 + c_2H(x) \ ,$$

for constants $c_1, c_2 \in \mathbb{F}$, where $H(x)$ is the Heaviside function. To see this, consider

$$xw = 0 \quad \text{in} \quad \mathcal{D}' \ .$$

For $\phi \in \mathcal{D}$,

$$0 = \langle xw, \phi\rangle = \langle w, x\phi\rangle \ ,$$

so we wish to write ϕ in terms of $x\psi$ for some $\psi \in \mathcal{D}$. To this end, let $r \in \mathcal{D}$ be any function that is 1 for $-\varepsilon < x < \varepsilon$ for some $\varepsilon > 0$ (such a function is easy to construct). Then

$$\begin{aligned}\phi(x) &= \phi(0)r(x) + (\phi(x) - \phi(0)r(x)) \\ &= \phi(0)r(x) + \int_0^x (\phi'(\xi) - \phi(0)r'(\xi))\, d\xi \\ &= \phi(0)r(x) + x\int_0^1 (\phi'(x\eta) - \phi(0)r'(x\eta))\, d\eta \\ &= \phi(0)r(x) + x\psi(x)\ ,\end{aligned}$$

where

$$\psi = \int_0^1 (\phi'(x\eta) - \phi(0)r'(x\eta))\, d\eta$$

clearly has compact support and $\psi \in C^\infty$, since differentiation and integration commute when the integrand is continuously differentiable. Thus

$$\langle w, \phi\rangle = \langle w, \phi(0)r\rangle + \langle w, x\psi\rangle = \phi(0)\langle w, r\rangle\ ;$$

that is, with $c = \langle w, r\rangle$,

$$w = c\delta_0\ .$$

Finally, then $v' = c_2\delta_0$ and $v = c_1 + c_2 H$. Our general solution

$$u = \ln|x| + c_1 + c_2 H(x)$$

is not a classical solution but merely a weak solution.

4.6.2. Partial Differential Equations and Fundamental Solutions. We return to $d \geq 1$ but restrict to the case of constant coefficients in L:

$$L = \sum_{|\alpha|\leq m} c_\alpha D^\alpha\ ,$$

where $c_\alpha \in \mathbb{F}$. We associate to L the polynomial

$$p(x) = \sum_{|\alpha|\leq m} c_\alpha x^\alpha\ ,$$

where $x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$; thus,

$$L = p(D)\ .$$

Easily, L is the adjoint of

$$\mathcal{L} = \sum_{|\alpha|\leq m} (-1)^{|\alpha|} c_\alpha D^\alpha\ ,$$

since $\langle u, \mathcal{L}\phi\rangle = \langle \mathcal{L}^* u, \phi\rangle = \langle Lu, \phi\rangle$ for any $u \in \mathcal{D}'$, $\phi \in \mathcal{D}$.

Example. Suppose L is the wave operator:

$$L = \frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2}$$

for $(t, x) \in \mathbb{R}^2$ and $c > 0$. For every $g \in C^2(\mathbb{R})$, $f(t, x) \equiv g(x - ct)$ solves $Lf = 0$. Similarly, if $g \in L_{1,\mathrm{loc}}$, we obtain a weak solution. In fact, $f(t, x) = \delta_0(x - ct)$ is a distributional solution, although we need to be more precise. Let $u \in \mathcal{D}'(\mathbb{R}^2)$ be defined by

$$\langle u, \phi\rangle = \langle \delta_0(x - ct), \phi(t, x)\rangle = \int_{-\infty}^{\infty} \phi(t, ct)\, dt \quad \forall\ \phi \in \mathcal{D}(\mathbb{R}^2)$$

(it is a simple exercise to verify that u is well defined in $\mathcal{D}'$). Then

$$\begin{aligned}\langle Lu,\phi\rangle &= \langle u,\mathcal{L}\phi\rangle = \langle u, L\phi\rangle\\ &= \Big\langle u, \Big(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2}\Big)\phi\Big\rangle\\ &= \Big\langle u, \Big(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\Big)\Big(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x}\Big)\phi\Big\rangle\\ &= \Big\langle u, \Big(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\Big)\psi\Big\rangle\ ,\end{aligned}$$

where $\psi \in \mathcal{D}$. Continuing,

$$\langle Lu,\phi\rangle = \int_{-\infty}^{\infty}\Big(\frac{\partial}{\partial t} + c\frac{\partial}{\partial x}\Big)\psi(t,ct)\,dt = \int_{-\infty}^{\infty}\frac{d}{dt}\psi(t,ct)\,dt = 0\ .$$

Definition. If $Lu = \delta_0$ for some $u \in \mathcal{D}'$, then u is called a *fundamental solution* of L.

If a fundamental solution u exists, it is *not* in general unique, since any solution to $Lv = 0$ gives another fundamental solution $u + v$. The reason for the name and its importance is given by the following theorem.

Theorem 4.26. *If $\psi \in \mathcal{D}$ and $u \in \mathcal{D}'$ is a fundamental solution for L, then $u*\psi$ is a solution to*

$$Lv = \psi\ .$$

Proof. Since $Lu = \delta_0$, then also

$$(Lu) * \psi = \delta_0 * \psi = \psi\ .$$

But

$$(Lu) * \psi = L(u * \psi)\ . \qquad \square$$

Theorem 4.27 (Malgrange and Ehrenpreis). *Every constant coefficient linear partial differential operator on $\mathbb{R}^d$ has a fundamental solution.*

A proof can be found in [**Yo**] and [**Ru1**].

Example. A fundamental solution of

$$L = \frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2}\ ,$$

where $c > 0$, is given by

$$u(t,x) = \frac{1}{2c}H(ct - |x|) = \frac{1}{2c}H(ct - x)H(ct + x)\ ,$$

where H is the Heaviside function. That is, we claim

$$\langle Lu,\phi\rangle = \phi(0,0)\quad \forall\ \phi \in \mathcal{D}'(\mathbb{R}^2)\ .$$

For convenience, let $D_\pm = \frac{\partial}{\partial t} \pm c\frac{\partial}{\partial x}$, so $L = D_+D_-$. Then

$$\begin{aligned}
\langle Lu, \phi\rangle = \langle u, D_+D_-\phi\rangle &= \iint \frac{1}{2c} H(ct-|x|) D_+D_-\phi\,dt\,dx \\
&= \frac{1}{2c}\Big\{\int_0^\infty \int_{x/c}^\infty D_+D_-\phi\,dt\,dx + \int_{-\infty}^0 \int_{-x/c}^\infty D_-D_+\phi\,dt\,dx\Big\} \\
&= \frac{1}{2c}\Big\{\int_0^\infty \int_0^\infty (D_+D_-\phi)(t+x/c,x)\,dt\,dx + \int_{-\infty}^0 \int_0^\infty (D_-D_+\phi)(t-x/c,x)\,dt\,dx\Big\} \\
&= \frac{1}{2}\Big\{\int_0^\infty \int_0^\infty \frac{d}{dx}(D_-\phi)(t+x/c,x)\,dx\,dt - \int_0^\infty \int_{-\infty}^0 \frac{d}{dx}(D_+\phi)(t-x/c,x)\,dx\,dt\Big\} \\
&= -\frac{1}{2}\int_0^\infty \Big[D_-\phi(t,0) + D_+\phi(t,0)\Big]\,dt \\
&= -\int_0^\infty \frac{\partial}{\partial t}\phi(t,0)\,dt \\
&= \phi(0,0) = \langle \delta_0, \phi\rangle \ .
\end{aligned}$$

EXAMPLE. The Laplace operator is

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_d^2} = \nabla\cdot\nabla = \nabla^2 \ .$$

A fundamental solution is given by

$$E(x) = \begin{cases} \dfrac{1}{2}|x| \ , & d=1 \ , \\ \dfrac{1}{2\pi}\ln|x| \ , & d=2 \ , \\ \dfrac{1}{d\omega_d}\,\dfrac{|x|^{2-d}}{2-d} \ , & d>2 \ , \end{cases} \tag{4.6}$$

where

$$\omega_d = \frac{2\pi^{d/2}}{d\Gamma(d/2)}$$

is the hyper-volume of the unit ball in $\mathbb{R}^d$. (As a side remark, the hyper-area of the unit sphere is $d\omega_d$.) It is trivial to verify the claim if $d=1$: $D^2\frac{1}{2}|x| = D\frac{1}{2}(2H(x)-1) = H' = \delta_0$. For $d \ge 2$, we need to show

$$\langle \Delta E, \phi\rangle = \langle E, \Delta\phi\rangle = \phi(0) \quad \forall\ \phi \in \mathcal{D}(\mathbb{R}^d) \ .$$

It is important to recognize that E is a regular distribution, i.e., $E \in L_{1,\mathrm{loc}}(\mathbb{R}^d)$. This is clear everywhere except possibly near $x=0$, where for $1 > r > 0$ and $d=2$, change of variables to polar coordinates gives

$$\begin{aligned}
\int_{B_r(0)} \Big|\frac{1}{2\pi}\ln|x|\Big|\,dx &= -\int_0^{2\pi}\int_0^r \frac{1}{2\pi}\ln r\, r dr\,d\theta \\
&= -\frac{1}{2}r^2\ln r + \frac{1}{4}r^2 < \infty
\end{aligned}$$

and, for $d > 2$,

$$\int_{B_r(0)} |E(x)|\,dx = -\int_{S_1(0)} \int_0^r \frac{r^{2-d}}{d\omega_d(2-d)} r^{d-1}\,dr\,d\sigma = \frac{r^2}{2(d-2)} < \infty \ ,$$

where $S_1(0)$ is the unit sphere. Thus we need that

$$\int E(x)\Delta\phi(x)\,dx = \phi(0) \quad \forall\ \phi \in \mathcal{D}\ .$$

Let $\operatorname{supp}(\phi) \subset B_R(0)$ and $\varepsilon > 0$. Then

$$\int_{\varepsilon<|x|<R} E\Delta\phi\,dx = -\int_{\varepsilon<|x|<R} \nabla E\cdot\nabla\phi\,dx + \int_{|x|=\varepsilon} E\nabla\phi\cdot\nu\,d\sigma\ ,$$

by the divergence theorem, where $\nu \in \mathbb{R}^d$ is the unit vector normal to the surface $|x| = \varepsilon$ pointing toward 0 (i.e., *out* of the set $\varepsilon < |x| < R$). Another application of the divergence theorem gives that

$$\int_{\varepsilon<|x|<R} E\Delta\phi\,dx = \int_{\varepsilon<|x|<R} \Delta E\phi\,dx - \int_{|x|=\varepsilon} \nabla E\cdot\nu\phi\,d\sigma + \int_{|x|=\varepsilon} E\nabla\phi\cdot\nu\,d\sigma\ .$$

It is an exercise to verify that $\Delta E = 0$ for $x \neq 0$. Moreover,

$$\int_{|x|=\varepsilon} E\nabla\phi\cdot\nu\,d\sigma = \int_{S_1(0)} \frac{1}{d\omega_d}\frac{\varepsilon^{2-d}}{2-d}\nabla\phi\cdot\nu\varepsilon^{d-1}\,d\sigma \to 0$$

as $\varepsilon \downarrow 0$ for $d > 2$ and similarly for $d = 2$. Also

$$\begin{aligned}-\int_{|x|=\varepsilon} \nabla E\cdot\nu\phi\,d\sigma &= \int_{S_1(0)} \frac{\partial E}{\partial r}(\varepsilon,\sigma)\phi(\varepsilon,\sigma)\varepsilon^{d-1}\,d\sigma \\ &= \int_{S_1(0)} \frac{1}{d\omega_d}\varepsilon^{1-d}\phi(\varepsilon,\sigma)\varepsilon^{d-1}\,d\sigma \longrightarrow \phi(0)\ .\end{aligned}$$

Thus

$$\int E\Delta\phi\,dx = \lim_{\varepsilon\downarrow 0}\int_{\varepsilon<|x|<R} E\Delta\phi\,dx = \phi(0)\ ,$$

as we needed to show.

If $f \in \mathcal{D}$, we can solve

$$\Delta u = f$$

by $u = E * f$. We can extend this result to many $f \in L_1$ by the following.

Theorem 4.28. *If $E(x)$ is the fundamental solution to the Laplacian given by* (4.6) *and $f \in L_1(\mathbb{R}^d)$ is such that for almost every $x \in \mathbb{R}^d$,*

$$E(x-y)f(y) \in L_1(\mathbb{R}^d)$$

(as a function of y), then

$$u = E * f$$

is well defined, $u \in L_{1,\mathrm{loc}}(\mathbb{R}^d)$, and

$$\Delta u = f \quad \text{in} \quad \mathcal{D}'\ .$$

PROOF. For any $r > 0$, using Fubini's theorem,

$$\int_{B_r(0)} |u(x)|\,dx \le \int_{B_r(0)} \int |E(x-y)f(y)|\,dy\,dx$$
$$= \int \int_{B_r(0)} |E(x-y)|\,dx |f(y)|\,dy < \infty\ ,$$

since $E \in L_{1,\mathrm{loc}}$ and $f \in L_1$. Thus $u \in L_{1,\mathrm{loc}}$.

For $\phi \in \mathcal{D}$, using again Fubini's theorem,

$$\begin{aligned}\langle \Delta u, \phi\rangle &= \langle u, \Delta\phi\rangle \\ &= \int u\Delta\phi\,dx = \iint E(x-y)f(y)\Delta\phi(x)\,dy\,dx \\ &= \iint E(x-y)\Delta\phi(x)\,dx\ f(y)\,dy \\ &= \int E * \Delta\phi(y)f(y)\,dy \\ &= \int \phi(y)f(y)\,dy = \langle f, \phi\rangle\ ,\end{aligned}$$

since $E(x-y) = E(y-x)$ and

$$E * \Delta\phi = \Delta E * \phi = \delta_0 * \phi = \phi\ .$$

Thus $\Delta u = f$ in $\mathcal{D}'$ as claimed. ☐

4.7. Local Structure of $\mathcal{D}'$

We state without proof the following theorem. See [**Ru1**, p. 154] for a proof.

THEOREM 4.29. *If $u \in \mathcal{D}'(\Omega)$, then there exist continuous functions g_α, one for each multi-index α, such that*

(i) *each $K \subset\subset \Omega$ intersects the supports of only finitely many of the g_α*

and

(ii) $u = \sum_\alpha D^\alpha g_\alpha$.

Thus we see that $\mathcal{D}'(\Omega)$ consists of nothing more than sums of derivatives of continuous functions, such that locally on any compact set, the sum is finite. Surely we wanted $\mathcal{D}'(\Omega)$ to contain at least all such functions. The complicated definition of $\mathcal{D}'(\Omega)$ we gave has included no other objects.

4.8. Exercises

1. Let $\psi \in \mathcal{D}$ be fixed and define $T : \mathcal{D} \to \mathcal{D}$ by $T(\phi) = \int \phi(\xi)\,d\xi\,\psi$. Show that T is a continuous linear map.
2. Show that if $\phi \in \mathcal{D}(\mathbb{R})$, then $\int \phi(x)\,dx = 0$ if and only if there is $\psi \in \mathcal{D}(\mathbb{R})$ such that $\phi = \psi'$.
3. Let T_h be the translation operator on $\mathcal{D}(\mathbb{R})$: $T_h\phi(x) = \phi(x-h)$. Show that for any $\phi \in \mathcal{D}(\mathbb{R})$,

$$\lim_{h\to 0} \frac{1}{h}(\phi - T_h\phi) = \phi' \quad \text{in } \mathcal{D}(\mathbb{R}).$$

4. Prove that $\mathcal{D}(\Omega)$ is *not* metrizable. [Hint: see the sketch of the proof given in Section 4.3.]
5. Prove directly that $x\,\mathrm{PV}(1/x) = 1$.
6. Let $T : \mathcal{D}(\mathbb{R}) \to \mathbb{R}$.
   (a) If $T(\phi) = |\phi(0)|$, show T is not a distribution.
   (b) If $T(\phi) = \sum_{n=0}^{\infty} \phi(n)$, show T is a distribution.
   (c) If $T(\phi) = \sum_{n=0}^{\infty} D^n\phi(n)$, show T is a distribution.
7. Is it true that $\delta_{1/n} \to \delta_0$ in $\mathcal{D}'$? Why or why not?
8. Determine if the following are distributions.
   (a) $\displaystyle\sum_{n=1}^{\infty} \delta_n = \lim_{N\to\infty} \sum_{n=1}^{N} \delta_n.$
   (b) $\displaystyle\sum_{n=1}^{\infty} \delta_{1/n} = \lim_{N\to\infty} \sum_{n=1}^{N} \delta_{1/n}.$
9. Let $\Omega \subset \mathbb{R}^d$ be open and let $\{a_n\}_{n=1}^{\infty}$ be a sequence from Ω with no accumulation point in Ω. For $\phi \in \mathcal{D}(\Omega)$, define
$$T(\phi) = \sum_{n=1}^{\infty} \lambda_n\,\phi(a_n),$$
where $\{\lambda_n\}_{n=1}^{\infty}$ is a sequence of complex numbers. Show that $T \in \mathcal{D}'(\Omega)$.
10. Prove the Plemelij-Sochozki formula $\dfrac{1}{x+i0} = \mathrm{PV}(1/x) - i\pi\delta_0(x)$; that is, for $\phi \in \mathcal{D}$,
$$\lim_{r\to 0}\left\{\lim_{\epsilon\to 0^+}\int_{|x|\geq\epsilon} \frac{1}{x+ir}\phi(x)\,dx\right\} = \lim_{\epsilon\to 0^+}\int_{|x|\geq\epsilon}\frac{1}{x}\phi(x)\,dx - i\pi\phi(0).$$
11. Prove that the trigonometric series $\displaystyle\sum_{n=-\infty}^{\infty} a_n e^{inx}$ converges in $\mathcal{D}'(\mathbb{R})$ if there exists a constant $A > 0$ and an integer $N \geq 0$ such that $|a_n| \leq A|n|^N$.
12. Show the following in $\mathcal{D}'(\mathbb{R})$.
    (a) $\displaystyle\lim_{n\to\infty} \cos(nx)\,\mathrm{PV}(1/x) = 0.$
    (b) $\displaystyle\lim_{n\to\infty} \sin(nx)\,\mathrm{PV}(1/x) = \pi\delta_0.$
    (c) $\displaystyle\lim_{n\to\infty} e^{inx}\,\mathrm{PV}(1/x) = i\pi\delta_0.$
13. Prove that the set of functions $\phi * \psi$, for ϕ and ψ in $\mathcal{D}$, is dense in $\mathcal{D}$.
14. Suppose that $u \in \mathcal{D}'$ and for any $\phi \in \mathcal{D}$, $u * \phi$ has compact support. For any $v \in \mathcal{D}'$, show that $v * (u * \phi)$ is well defined. Further define $v * u$, show that it is in $\mathcal{D}'$, and that $(v * u) * \phi = v * (u * \phi)$.
15. Find a general solution to the differential equation $D^2T = 0$ in $\mathcal{D}'(\mathbb{R})$.
16. Verify that $\Delta E = 0$ for $x \neq 0$, where E is the fundamental solution to the Laplacian given in the text.

17. Find a fundamental solution for the operator $-D^2 + I$ on $\mathbb{R}$.

18. On $\mathbb{R}^3$, show that the operator

$$T(\phi) = \lim_{\epsilon \to 0^+} \int_{|x| \geq \epsilon} \frac{1}{4\pi|x|} e^{-|kx|} \, \phi(x) \, dx$$

is a fundamental solution to the Helmholtz operator $-\Delta + k^2 I$.

CHAPTER 5

The Fourier Transform

Fourier analysis began with Jean-Baptiste-Joseph Fourier's work two centuries ago. Fourier was concerned with the propagation of heat and invented what we now call Fourier series. He used a Fourier series representation to express solutions of the linear heat equation. His work was greeted with suspicion by his contemporaries.

The paradigm that Fourier put forward has proved to be a central conception in analysis and in the theory of differential equations. The idea is this. Consider for example the linear heat equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \dfrac{\partial^2 u}{\partial x^2} \ , \qquad 0 < x < 1 \ , \ t > 0 \ , \\ u(0,t) = u(1,t) = 0 \ , \\ u(x,0) = \varphi(x) \ , \end{cases} \tag{5.1}$$

in which the ends of the bar are held at constant temperature 0, and the initial temperature distribution $\varphi(x)$ is given. This might look difficult to solve, so let us try a special case

$$\varphi(x) = \sin(n\pi x) \ , \qquad n = 1, 2, \dots \ .$$

Try for a solution of the form

$$u_n(x,t) = U_n(t) \sin(n\pi x) \ .$$

Then U_n has to satisfy

$$U_n' \sin(n\pi x) = -n^2 U_n \sin(n\pi x) \ ,$$

or

$$U_n' = -n^2 U_n \ . \tag{5.2}$$

We can solve this very easily:

$$U_n(t) = U_n(0) e^{-n^2 t} \ .$$

The solution is

$$u_n(x,t) = U_n(0) e^{-n^2 t} \sin(n\pi x) \ .$$

Now, and here is Fourier's great conception, suppose we can decompose φ into $\{\sin(n\pi x)\}_{n=1}^{\infty}$:

$$\varphi(x) = \sum_{n=1}^{\infty} \varphi_n \sin(n\pi x) \ ;$$

that is, we represent φ in terms of the simple functions $\sin(n\pi x)$, $n = 1, 2, \ldots$. Then we obtain formally a representation of the solution of (5.1), namely

$$u(x,t) = \sum_{n=1}^{\infty} u_n(x,t) = \sum_{n=1}^{\infty} \varphi_n e^{-n^2 t} \sin(n\pi x) \ .$$

In obtaining this, we used the representation in terms of simple harmonic functions $\{\sin(n\pi x)\}_{n=1}^{\infty}$ to convert the partial differential equation (PDE) (5.1) into a system of ordinary differential equations (ODE's) (5.2).

Suppose now the rod was infinitely long, so we want to solve

$$\begin{cases} u_t = u_{xx} \ , \quad -\infty < x < \infty \ , \ t > 0 \ , \\ u(x,0) = \varphi(x) \ , \\ u(x,t) \to 0 \ \text{ as } \ x \to \pm\infty \ . \end{cases} \tag{5.3}$$

Again, we would like to represent φ in terms of harmonic functions, e.g.,

$$\varphi(x) = \sum_{n=-\infty}^{\infty} \varphi_n e^{-inx} \ .$$

Any such function is periodic of period 2π, however. It turns out that to represent a general function, you need the uncountable class

$$\{e^{-i\lambda x}\}_{\lambda \in \mathbb{R}} \ .$$

We cannot sum these, but we might be able to integrate them; viz.,

$$\varphi(x) = \int_{-\infty}^{\infty} e^{-i\lambda x} \rho(\lambda) \, d\lambda \ ,$$

say for some density ρ. Suppose we could. As before, we search for a solution in the form

$$U(x,t) = \int_{-\infty}^{\infty} e^{-i\lambda x} \rho(\lambda, t) \, d\lambda \ .$$

If this is to satisfy (5.3), then

$$\int_{-\infty}^{\infty} e^{-i\lambda x} \frac{\partial \rho}{\partial t}(\lambda, t) \, d\lambda = -\int_{-\infty}^{\infty} e^{-i\lambda x} \lambda^2 \rho(\lambda, t) \, d\lambda \ ,$$

or

$$\int_{-\infty}^{\infty} e^{-i\lambda x} \left[\frac{\partial \rho}{\partial t} + \lambda^2 \rho \right] d\lambda = 0 \ ,$$

for all x, t. As x is allowed to wonder over all of $\mathbb{R}$, we conclude that this will hold only when

$$\frac{\partial \rho}{\partial t} + \lambda^2 \rho = 0 \qquad \forall \ \lambda \in \mathbb{R} \ . \tag{5.4}$$

This collection of ODE's is easily solved as before:

$$\rho(\lambda, t) = \rho(\lambda, 0) e^{-\lambda^2 t} \ .$$

Thus formally, the full solution is

$$u(x,t) = \int_{-\infty}^{\infty} e^{-i\lambda x} e^{-\lambda^2 t} \rho(\lambda) \, d\lambda \ ,$$

another representation of solutions. These observations that

(1) functions can be represented in terms of harmonic functions, and
(2) in this representation, PDE's may be reduced in complexity to ODE's,

is already enough to warrant further study. The crux of the formula above for u is ρ — what is ρ such that

$$\varphi(x) = \int_{-\infty}^{\infty} e^{-i\lambda x}\rho(\lambda)\,d\lambda\ ?$$

Is there such a ρ, and if so, how do we find it? This leads us directly to the study of the Fourier transform: $\mathcal{F}$.

The Fourier transform is a linear operator that can be defined naturally for any function in $L_1(\mathbb{R}^d)$. The definition can be extended to apply to functions in $L_2(\mathbb{R}^d)$, and then the transform takes $L_2(\mathbb{R}^d)$ onto itself with nice properties. Moreover, the Fourier transform can be applied to some, but unfortunately not all, distributions, called tempered distributions.

Throughout this chapter we assume that the underlying vector space field $\mathbb{F}$ is $\mathbb{C}$.

5.1. The $L_1(\mathbb{R}^d)$ theory

If $\xi \in \mathbb{R}^d$, the function

$$\varphi_\xi(x) = e^{-ix\cdot\xi} = \cos(x\cdot\xi) - i\sin(x\cdot\xi)\ , \qquad x \in \mathbb{R}^d\ ,$$

is a wave in the direction ξ. Its period in the j^{th} direction is $2\pi/\xi_j$. These functions have nice algebraic and differential properties.

Proposition 5.1.

(a) $|\varphi_\xi| = 1$ *and* $\bar\varphi_\xi = \varphi_{-\xi}$ *for any* $\xi \in \mathbb{R}^d$.
(b) $\varphi_\xi(x+y) = \varphi_\xi(x)\varphi_\xi(y)$ *for any* $x, y, \xi \in \mathbb{R}^d$.
(c) $-\Delta\varphi_\xi = |\xi|^2\varphi_\xi$ *for any* $\xi \in \mathbb{R}^d$.

These are easily verified. Note that the third result says that φ_ξ is an eigenfunction of the Laplace operator with eigenvalue $-|\xi|^2$.

If $f(x)$ is periodic, we can expand f as a Fourier series using commensurate waves $e^{-ix\cdot\xi}$ (i.e., waves of the same period) as mentioned above. If f is not periodic, we need all such waves. This leads us to the Fourier transform, which has nice algebraic and differential properties similar to those listed above for $e^{-ix\cdot\xi}$.

Definition. If $f \in L_1(\mathbb{R}^d)$, the *Fourier transform* of f is

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

This is well defined since

$$|f(x)e^{-ix\cdot\xi}| = |f(x)| \in L_1(\mathbb{R}^d)\ .$$

We remark that it is possible to define a Fourier transform by any of the following:

$$\int_{\mathbb{R}^d} f(x)e^{\pm 2\pi i x\cdot\xi}\,dx\ ,$$

$$\int_{\mathbb{R}^d} f(x)e^{\pm i x\cdot\xi}\,dx\ ,$$

$$(2\pi)^{-d/2}\int_{\mathbb{R}^d} f(x)e^{\pm i x\cdot\xi}\,dx\ .$$

The choice here affects the form of the results that follow, but not their substance. Different authors make different choices here, but it is easy to translate results for one definition into another.

PROPOSITION 5.2. *The Fourier transform*

$$\mathcal{F} : L_1(\mathbb{R}^d) \to L_\infty(\mathbb{R}^d)$$

is a bounded linear operator, and

$$\|\hat{f}\|_{L_\infty(\mathbb{R}^d)} \le (2\pi)^{-d/2} \|f\|_{L_1(\mathbb{R}^d)} \ .$$

The proof is an easy exercise of the definitions.

EXAMPLE. Consider the characteristic function of $[-1,1]^d$:

$$f(x) = \begin{cases} 1 & \text{if } -1 < x_j < 1,\ j = 1, \dots, d, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} \hat{f}(\xi) &= (2\pi)^{-d/2} \int_{-1}^{1} \cdots \int_{-1}^{1} e^{-ix\cdot\xi}\, dx \\ &= \prod_{j=1}^{d} (2\pi)^{-1/2} \int_{-1}^{1} e^{-ix_j\xi_j}\, dx_j \\ &= \prod_{j=1}^{d} (2\pi)^{-1/2} \frac{-1}{i\xi_j} (e^{-i\xi_j} - e^{i\xi_j}) \\ &= \prod_{j=1}^{d} \sqrt{\frac{2}{\pi}}\, \frac{\sin \xi_j}{\xi_j} \ . \end{aligned}$$

PROPOSITION 5.3. *If $f \in L_1(\mathbb{R}^d)$ and τ_y is translation by y (i.e., $\tau_y\varphi(x) = \varphi(x-y)$), then*

(a) $(\tau_y f)^\wedge(\xi) = e^{-iy\cdot\xi}\hat{f}(\xi) \quad \forall\ y \in \mathbb{R}^d$;
(b) $(e^{ix\cdot y} f)^\wedge(\xi) = \tau_y \hat{f}(\xi) \quad \forall\ y \in \mathbb{R}^d$;
(c) *if $r > 0$ is given,*

$$\widehat{f(rx)}(\xi) = r^{-d}\hat{f}(r^{-1}\xi) \ ;$$

(d) $\hat{\hat{f}}(\xi) = \overline{\hat{f}(-\xi)}$.

The proof is a simple exercise of change of variables.

While the Fourier transform maps $L_1(\mathbb{R}^d)$ *into* $L_\infty(\mathbb{R}^d)$, it does not map *onto*. Its range is poorly understood, but it is known to be contained in a set we will call $C_v(\mathbb{R}^d)$.

DEFINITION. A continuous function f on $\mathbb{R}^d$ is said to *vanish at infinity* if for any $\varepsilon > 0$ there is $K \subset\subset \mathbb{R}^d$ such that

$$|f(x)| < \varepsilon \quad \forall\ x \notin K \ .$$

We define

$$C_v(\mathbb{R}^d) = \{ f \in C^0(\mathbb{R}^d) : f \text{ vanishes at } \infty \} \ .$$

PROPOSITION 5.4. *The space $C_v(\mathbb{R}^d)$ is a closed linear subspace of $L_\infty(\mathbb{R}^d)$.*

PROOF. Suppose that $\{f_n\}_{n=1}^\infty \subset C_v(\mathbb{R}^d)$ and that

$$f_n \xrightarrow{L_\infty} f \ .$$

Then f is continuous (the uniform convergence of continuous functions is continuous). Now let $\varepsilon > 0$ be given and choose n such that $\|f - f_n\|_{L_\infty} < \varepsilon/2$ and $K \subset\subset \mathbb{R}^d$ such that $|f_n(x)| < \varepsilon/2$ for $x \notin K$. Then

$$|f(x)| \leq |f(x) - f_n(x)| + |f_n(x)| < \varepsilon$$

shows that $f \in C_v(\mathbb{R}^d)$. □

LEMMA 5.5 (Riemann-Lebesgue Lemma). *The Fourier transform*

$$\mathcal{F} : L_1(\mathbb{R}^d) \to C_v(\mathbb{R}^d) \subsetneqq L_\infty(\mathbb{R}^d) \ .$$

Thus for $f \in L_1(\mathbb{R}^d)$,

$$\lim_{|\xi|\to\infty} |\hat{f}(\xi)| = 0 \quad \text{and} \quad \hat{f} \in C^0(\mathbb{R}^d) \ .$$

PROOF. Let $f \in L_1(\mathbb{R}^d)$. There is a sequence of simple functions $\{f_n\}_{n=1}^\infty$ such that $f_n \to f$ in $L_1(\mathbb{R}^d)$. Recall that a simple function is a finite linear combination of characteristic functions of rectangles. If $\hat{f}_n \in C_v(\mathbb{R}^d)$, we are done since

$$\hat{f}_n \xrightarrow{L_\infty} \hat{f}$$

and $C_v(\mathbb{R}^d)$ is a closed subspace. We know that the Fourier transform of the characteristic function of $[-1,1]^d$ is

$$\prod_{j=1}^d \sqrt{\frac{2}{\pi}} \, \frac{\sin \xi_j}{\xi_j} \in C_v(\mathbb{R}^d) \ .$$

By Proposition 5.3, translation and dilation of this cube gives us that the characteristic function of any rectangle is in $C_v(\mathbb{R}^d)$, and hence also any finite linear combination of these. □

Some nice properties of the Fourier transform are given in the following.

PROPOSITION 5.6. *If* $f, g \in L_1(\mathbb{R}^d)$, *then*

(a) $\displaystyle\int \hat{f}(x) g(x)\, dx = \int f(x) \hat{g}(x)\, dx$,

(b) $f * g \in L_1(\mathbb{R}^d)$ *and* $\widehat{f * g} = (2\pi)^{d/2} \hat{f}\hat{g}$,

where

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

is defined for almost every $x \in \mathbb{R}^d$.

Proof. For (a), note that $\hat{f} \in L_\infty$ and $g \in L_1$ implies $\hat{f}g \in L_1$, so the integrals are well defined. Fubini's theorem gives the result:

$$\begin{aligned}\int \hat{f}(x)g(x)\,dx &= (2\pi)^{-d/2}\iint f(y)e^{-ix\cdot y}g(x)\,dy\,dx\\ &= (2\pi)^{-d/2}\iint f(y)e^{-ix\cdot y}g(x)\,dx\,dy\\ &= \int f(y)\hat{g}(y)\,dy\ .\end{aligned}$$

The reader can show (b) similarly, using Fubini's theorem and change of variables, once we know that $f * g \in L_1(\mathbb{R}^d)$. We show this fact below, more generally than we need here. ☐

Theorem 5.7 (Generalized Young's Inequality). *Suppose $K(x, y)$ is measurable on $\mathbb{R}^d \times \mathbb{R}^d$ and there is some $C > 0$ such that*

$$\int |K(x,y)|\,dx \le C \quad \textit{for almost every} \quad y \in \mathbb{R}^d$$

and

$$\int |K(x,y)|\,dy \le C \quad \textit{for almost every} \quad x \in \mathbb{R}^d\ .$$

Let the operator T be defined by

$$Tf(x) = \int K(x,y)f(y)\,dy\ .$$

If $1 \le p \le \infty$, then $T : L_p(\mathbb{R}^d) \to L_p(\mathbb{R}^d)$ is a bounded linear map with norm $\|T\| \le C$.

Corollary 5.8 (Young's Inequality). *If $1 \le p \le \infty$, $f \in L_p(\mathbb{R}^d)$, and $g \in L_1(\mathbb{R}^d)$, then $f * g \in L_p(\mathbb{R}^d)$ and*

$$\|f * g\|_{L_p(\mathbb{R}^d)} \le \|f\|_{L_p(\mathbb{R}^d)}\|g\|_{L_1(\mathbb{R}^d)}\ .$$

Just take $K(x, y) = g(x - y)$.

Corollary 5.9. *The space $L_1(\mathbb{R}^d)$ is an algebra with multiplication defined by the convolution operation.*

Proof. (Generalized Young's Inequality) If $p = \infty$, the result is trivial (and, in fact, we need not assume that $\int |K(x, y)|\,dx \le C$). If $p < \infty$, let $\frac{1}{q} + \frac{1}{p} = 1$ and then

$$\begin{aligned}|Tf(x)| &\le \int |K(x,y)|^{1/q}|K(x,y)|^{1/p}|f(y)|\,dy\\ &\le \left(\int |K(x,y)|\,dy\right)^{1/q}\left(\int |K(x,y)|\,|f(y)|^p\,dy\right)^{1/p}\end{aligned}$$

by Hölder's inequality. Thus

$$\begin{aligned}\|Tf\|_{L_p}^p &\leq C^{p/q} \iint |K(x,y)|\,|f(y)|^p\,dy\,dx \\ &= C^{p/q} \iint |K(x,y)|\,dx|f(y)|^p\,dy \\ &\leq C^{p/q+1} \int |f(y)|^p\,dy \\ &= C^p \|f\|_{L_p}^p\ ,\end{aligned}$$

and the theorem follows since T is clearly linear. $\square$

An unresolved question is: Given f, what does $\hat{f}$ look like? We have the Riemann-Lebesgue lemma, and the following theorem.

THEOREM 5.10 (Paley-Wiener). *If $f \in C_0^\infty(\mathbb{R}^d)$, then $\hat{f}$ extends to an entire holomorphic function on $\mathbb{C}^d$.*

PROOF. The function

$$\xi \longmapsto e^{-ix\cdot\xi}$$

is an entire function for $x \in \mathbb{R}^d$ fixed. The Riemann sums approximating

$$\hat{f}(\xi) = (2\pi)^{-d/2} \int f(x) e^{-ix\cdot\xi}\,dx$$

are entire, and they converge uniformly on compact sets since $f \in C_0^\infty(\mathbb{R}^d)$. Thus we conclude that $\hat{f}$ is entire. $\square$

See [**Ru1**] for the converse. Since holomorphic functions do *not* have compact support, we see that functions which are localized in space are *not* localized in Fourier space (and conversely).

5.2. The Schwartz space theory

Since $L_2(\mathbb{R}^d)$ is not contained in $L_1(\mathbb{R}^d)$, we restrict to a suitable subspace $\mathcal{S} \subset L_2(\mathbb{R}^d) \cap L_1(\mathbb{R}^d)$ on which to define the Fourier transform before attempting the definition on $L_2(\mathbb{R}^d)$.

DEFINITION. The *Schwartz space* or *space of functions of rapid decrease* is

$$\mathcal{S} = \mathcal{S}(\mathbb{R}^d) = \left\{\phi \in C^\infty(\mathbb{R}^d) : \sup_{x\in\mathbb{R}^d} |x^\alpha D^\beta \phi(x)| < \infty \text{ for all multi-indices } \alpha \text{ and } \beta\right\}\ .$$

That is, ϕ and all its derivatives tend to 0 at infinity faster than any polynomial. As an example, consider $\phi(x) = p(x)e^{-a|x|^2}$ for any $a > 0$ and any polynomial $p(x)$.

PROPOSITION 5.11. *One has that*

$$C_0^\infty(\mathbb{R}^d) \subsetneqq \mathcal{S} \subsetneqq L_1(\mathbb{R}^d) \cap L_\infty(\mathbb{R}^d)\ ;$$

thus also $\mathcal{S}(\mathbb{R}^d) \subset L_p(\mathbb{R}^d)\ \forall\ 1 \leq p \leq \infty$.

PROOF. The only nontrivial statement is that $\mathcal{S} \subset L_1$. For $\phi \in \mathcal{S}$,

$$\int |\phi(x)|\,dx = \int_{B_1(0)} |\phi(x)|\,dx + \int_{|x|\geq 1} |\phi(x)|\,dx\ .$$

The former integral is finite, so consider the latter. Since $\phi \in \mathcal{S}$, we can find $C > 0$ such that $|x|^{d+1}|\phi(x)| < C \ \forall \ |x| > 1$. Then

$$\begin{aligned}\int_{|x|\geq 1} |\phi(x)|\,dx &= \int_{|x|\geq 1} |x|^{-d-1}(|x|^{d+1}|\phi(x)|)\,dx \\ &\leq C \int_{|x|\geq 1} |x|^{-d-1}\,dx \\ &\leq C\,d\omega_d \int_1^\infty r^{-d-1} r^{d-1}\,dr \\ &= C\,d\omega_d \int_1^\infty r^{-2}\,dr < \infty \ ,\end{aligned}$$

where $d\omega_d$ is the measure of the unit sphere. □

Given $n = 0, 1, 2, \ldots$, we define for $\phi \in \mathcal{S}$

$$\rho_n(\phi) = \sup_{|\alpha|\leq n} \sup_x (1 + |x|^2)^{n/2} |D^\alpha \phi(x)| \ . \tag{5.5}$$

Each ρ_n is a norm on $\mathcal{S}$ and $\rho_n(\phi) \leq \rho_m(\phi)$ whenever $n \leq m$.

Proposition 5.12. *The Schwartz class $\mathcal{S} = \{\phi \in C^\infty : \rho_n(\phi) < \infty \ \forall \ n\}$, and $\mathcal{S}$ is a complete metric space where the $\{\rho_n\}_{n=0}^\infty$ generate its topology through the metric*

$$d(\phi_1, \phi_2) = \sum_{n=0}^\infty 2^{-n} \frac{\rho_n(\phi_1 - \phi_2)}{1 + \rho_n(\phi_1 - \phi_2)} \ .$$

We remark that for a sequence in $\mathcal{S}$, $\phi_j \to \phi$ if and only if $\rho_n(\phi_1 - \phi_2) \to 0$ for all n.

Proof. Clearly $\mathcal{S}$ is a vector space and d is a metric. Also

$$\mathcal{S} = \{\phi \in C^\infty : \rho_n(\phi) < \infty \ \forall \ n\} \ ,$$

because sums of terms like $\omega_{\alpha\beta}(\phi) = \sup_x |x^\alpha D^\beta \phi|$ bound $\rho_n(\phi)$, and $\rho_n(\phi)$ bounds $\omega_{\alpha\beta}(\phi)$ for $n = \max(|\alpha|, |\beta|)$.

It remains to show completeness. Let $\{\phi_j\}_{j=1}^\infty$ be a Cauchy sequence in $\mathcal{S}$. That is,

$$\rho_n(\phi_j - \phi_k) \to 0 \ \text{ as } \ j, k \to \infty \ \forall \ n \ .$$

Thus, for any α and $n \geq |\alpha|$,

$$\left\{(1 + |x|^2)^{n/2} D^\alpha \phi_j\right\}_{j=1}^\infty \ \text{ is Cauchy in } \ C^0(\mathbb{R}^d) \ ,$$

so there is some $\psi_{n,\alpha} \in C^0(\mathbb{R}^d)$ such that

$$(1 + |x|^2)^{n/2} D^\alpha \phi_j \xrightarrow{L_\infty} \psi_{n,\alpha} \ .$$

But then it follows

$$D^\alpha \phi_j \xrightarrow{L_\infty} \frac{\psi_{n,\alpha}}{(1 + |x|^2)^{n/2}} \in C^0(\mathbb{R}^d) \ .$$

Now $\phi_j \xrightarrow{L_\infty} \psi_{0,0}$, so as distributions $D^\alpha \phi_j \xrightarrow{\mathcal{D}'} D^\alpha \psi_{0,0}$. So $\psi_{n,\alpha} = (1 + |x|^2)^{n/2} D^\alpha \psi_{0,0}$, $\rho_n(\psi_{0,0}) < \infty \ \forall \ n$, and $\rho_n(\phi_j - \psi_{0,0}) \to 0 \ \forall \ n$. That is, $\psi_{0,0} \in \mathcal{S}$, and $\phi_j \xrightarrow{\mathcal{S}} \psi_{0,0}$. □

PROPOSITION 5.13. *If $p(x)$ is a polynomial, $g \in \mathcal{S}$, and α a multi-index, then each of the three mappings*

$$f \mapsto pf \ , \quad f \mapsto gf \ , \quad \text{and} \quad f \mapsto D^\alpha f$$

is a continuous linear map from $\mathcal{S}$ to $\mathcal{S}$.

PROOF. The range of each map is $\mathcal{S}$, by the Leibniz formula for the first two. Each map is easily seen to be sequentially continuous, thus continuous. □

Since $\mathcal{S} \subset L_1(\mathbb{R}^d)$, we can take the Fourier transform of functions in $\mathcal{S}$.

THEOREM 5.14. *If $f \in \mathcal{S}$ and α is a multi-index, then*

(a) $(D^\alpha f)^\wedge(\xi) = (i\xi)^\alpha \hat{f}(\xi)$,
(b) $D^\alpha \hat{f}(\xi) = \big((-ix)^\alpha f(x)\big)^\wedge(\xi)$.

PROOF. For (a)

$$\begin{aligned}
(2\pi)^{d/2}(D^\alpha f)^\wedge(\xi) &= \int D^\alpha f(x) e^{-ix\cdot\xi}\, dx \\
&= \lim_{r\to\infty} \int_{B_r(0)} D^\alpha f(x) e^{-ix\cdot\xi}\, dx \\
&= \lim_{r\to\infty} \left\{ \int_{B_r(0)} f(x)(i\xi)^\alpha e^{-ix\cdot\xi}\, dx + (\text{boundary terms}) \right\} ,
\end{aligned}$$

by integration by parts. There are finitely many boundary terms, each evaluated at $|x| = r$ and the absolute value of any such boundary term is bounded by a constant times $|D^\beta f(x)|$ for some multi-index $\beta \le \alpha$. Since $f \in \mathcal{S}$, each of these tends to zero faster than the measure of $\partial B_r(0)$ (i.e., faster than r^{d-1}), so each boundary term vanishes. Continuing,

$$(D^\alpha f)^\wedge(\xi) = (2\pi)^{-d/2} \int f(x)(i\xi)^\alpha e^{-ix\cdot\xi}\, dx = (i\xi)^\alpha \hat{f}(\xi) \ .$$

For (b), we wish to interchange integration and differentiation, since

$$(2\pi)^{d/2} D^\alpha \hat{f}(\xi) = D^\alpha \int f(x) e^{-ix\cdot\xi}\, dx \ .$$

Consider a single derivative

$$(2\pi)^{d/2} D_j \hat{f}(\xi) = \lim_{h\to 0} \int f(x) e^{-ix\cdot\xi} \frac{e^{-ix_j h} - 1}{h}\, dx \ .$$

Since

$$\left|\frac{e^{-i\theta} - 1}{\theta}\right|^2 = 2\left|\frac{1 - \cos\theta}{\theta^2}\right| \le 1 \ ,$$

we have

$$\left| ix_j f(x) e^{-ix\cdot\xi} \frac{e^{-ix_j h} - 1}{ix_j h} \right| \le |x_j f(x)| \in L_1$$

independently of h, and the Dominated Convergence theorem applies and shows that

$$\begin{aligned}(2\pi)^{d/2} D_j \hat{f}(\xi) &= \int \lim_{h\to 0} i x_j f(x)\ e^{-ix\cdot\xi} \frac{e^{-ix_j h}-1}{ix_j h}\, dx \\ &= \int -ix_j f(x) e^{-ix\cdot\xi}\, dx \\ &= (2\pi)^{d/2}\big(-ix_j f(x)\big)^{\wedge}(\xi)\ .\end{aligned}$$

By iteration, we obtain the result for $D^\alpha \hat{f}$. □

LEMMA 5.15. *The Fourier transform $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is continuous and linear.*

PROOF. We first show that the range is $\mathcal{S}$. For $f \in \mathcal{S}$, $x^\alpha D^\beta f \in L_\infty$ for any multi-indices α and β. But then

$$\xi^\alpha D^\beta \hat{f} = \xi^\alpha \big((-ix)^\beta f\big)^{\wedge} = (-1)^{|\beta|} i^{|\beta|-|\alpha|} \big(D^\alpha(x^\beta f)\big)^{\wedge}\ ,$$

and so

$$\|\xi^\alpha D^\beta \hat{f}\|_{L_\infty} \le (2\pi)^{-d/2} \|D^\alpha(x^\beta f)\|_{L_1} < \infty\ ,$$

since $D^\alpha(x^\beta f)$ rapidly decreases, and we conclude that $\hat{f} \in \mathcal{S}$.

The linearity of $\mathcal{F}$ is clear. Now if $\{f_j\}_{j=1}^\infty \subset \mathcal{S}$ and $f_j \xrightarrow{\mathcal{S}} f$, then also $f_j \xrightarrow{L_1} f$. Since $\mathcal{F}$ is continuous on L_1, $\hat{f}_j \xrightarrow{L_\infty} \hat{f}$. Similarly we conclude

$$(x^\alpha D^\beta f_j)^{\wedge} \xrightarrow{L_\infty} (x^\alpha D^\beta f)^{\wedge}\ ,$$

and thus that $\hat{f}_j \xrightarrow{\mathcal{S}} f$. □

In fact, after the following lemma, we show that $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is one-to-one and maps onto $\mathcal{S}$.

LEMMA 5.16. *If $\phi(x) = e^{-|x|^2/2}$, then $\phi \in \mathcal{S}$ and $\hat{\phi}(\xi) = \phi(\xi)$.*

PROOF. The reader can easily verify that $\phi \in \mathcal{S}$. Since

$$\begin{aligned}\hat{\phi}(\xi) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{-|x|^2/2} e^{-ix\cdot\xi}\, dx \\ &= \prod_{j=1}^d (2\pi)^{-1/2} \int_{\mathbb{R}} e^{-x_j^2/2} e^{-ix_j\xi_j}\, dx_j\ ,\end{aligned}$$

we need only show the result for $d = 1$. This can be accomplished directly using complex contour integration and Cauchy's Theorem. An alternate proof is to note that for $d = 1$, $\phi(x)$ solves

$$y' + xy = 0$$

and $\hat{\phi}(\xi)$ solves

$$0 = \widehat{y'} + \widehat{xy} = i\xi\hat{y} + i\hat{y}'\ ,$$

the same equation. Thus $\hat{\phi}/\phi$ is constant. But $\phi(0) = 1$ and $\hat{\phi}(0) = (2\pi)^{-1/2}\int e^{-x^2/2}\, dx = 1$, so $\hat{\phi} = \phi$. □

THEOREM 5.17. *The Fourier transform $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is a continuous, linear, one-to-one map of $\mathcal{S}$ onto $\mathcal{S}$ with a continuous inverse. The map $\mathcal{F}$ has period 4, and in fact $\mathcal{F}^2$ is reflection about the origin. If $f \in \mathcal{S}$, then*

$$f(x) = (2\pi)^{-d/2} \int \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi \ . \tag{5.6}$$

Moreover, if $f \in L_1(\mathbb{R}^d)$ and $\hat{f} \in L_1(\mathbb{R}^d)$, then (5.6) holds for almost every $x \in \mathbb{R}^d$.

Sometimes we write $\mathcal{F}^{-1} = \check{\bullet}$ for the inverse Fourier transform:

$$\mathcal{F}^{-1}(g)(x) = \check{g}(x) = (2\pi)^{-d/2} \int g(\xi) e^{ix\cdot\xi}\, d\xi \ .$$

PROOF. We first prove (5.6) for $f \in \mathcal{S}$. Let $\phi \in \mathcal{S}$ and $\varepsilon > 0$. Then

$$\int f(x) \varepsilon^{-d} \hat{\phi}(\varepsilon^{-1} x)\, dx = \int f(\varepsilon y) \hat{\phi}(y)\, dy \to f(0) \int \hat{\phi}(y)\, dy$$

as $\varepsilon \to 0$ by the Dominated Convergence Theorem since $f(\varepsilon y) \to f(0)$ uniformly. (We have just shown that $\varepsilon^{-d}\hat{\phi}(\varepsilon^{-1}x)$ converges to a multiple of δ_0 in $\mathcal{S}'$.) But also

$$\int f(x) \varepsilon^{-d} \hat{\phi}(\varepsilon^{-1} x)\, dx = \int \hat{f}(x) \phi(\varepsilon x)\, dx \to \phi(0) \int \hat{f}(x)\, dx \ ,$$

so

$$f(0) \int \hat{\phi}(y)\, dy = \phi(0) \int \hat{f}(x)\, dx \ .$$

Take

$$\phi(x) = e^{-|x|^2/2} \in \mathcal{S}$$

to see by the lemma that

$$f(0) = (2\pi)^{-d/2} \int \hat{f}(\xi)\, d\xi \ ,$$

which is (5.6) for $x = 0$. The general result follows by translation:

$$\begin{aligned} f(x) &\equiv (\tau_{-x} f)(0) \\ &= (2\pi)^{-d/2} \int (\tau_{-x} f)^{\wedge}(\xi)\, d\xi \\ &= (2\pi)^{-d/2} \int e^{ix\cdot\xi} \hat{f}(\xi)\, d\xi \ . \end{aligned}$$

We saw earlier that $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is continuous and linear; it is one-to-one by (5.6). Moreover,

$$\mathcal{F}^2 f(x) = f(-x)$$

follows as a simple computation since $\mathcal{F}$ and $\mathcal{F}^{-1}$ are so similar. Thus $\mathcal{F}$ maps onto $\mathcal{S}$, $\mathcal{F}^4 = I$, $\mathcal{F}^{-1} = \mathcal{F}^3$ is continuous.

It remains to extend (5.6) to $L_1(\mathbb{R}^d)$. If $f, \hat{f} \in L_1(\mathbb{R}^d)$, then we can define

$$f_0(x) = \check{\hat{f}}(x) = (2\pi)^{-d/2} \int \hat{f}(\xi) e^{ix\cdot\xi}\, d\xi \ .$$

Then for $\phi \in \mathcal{S}$,

$$\begin{aligned}\int f(x)\hat{\phi}(x)\,dx &= \int \hat{f}(x)\phi(x)\,dx \\ &= (2\pi)^{-d/2}\int \hat{f}(x)\int \hat{\phi}(\xi)e^{ix\cdot\xi}\,d\xi\,dx \\ &= (2\pi)^{-d/2}\iint \hat{f}(x)e^{ix\cdot\xi}\hat{\phi}(\xi)\,dx\,d\xi \\ &= \int f_0(\xi)\hat{\phi}(\xi)\,d\xi\ ,\end{aligned}$$

and we conclude by the Lebesgue Lemma that

$$f(x) = f_0(x)$$

for almost every $x \in \mathbb{R}^d$, since $\hat{\phi}(x)$ is an arbitrary member of $\mathcal{S}$ (since $\mathcal{F}$ maps onto). $\square$

We conclude the $\mathcal{S}$ theory with a result about convolutions.

THEOREM 5.18. *If $f, g \in \mathcal{S}$, then $f * g \in \mathcal{S}$ and*

$$(2\pi)^{d/2}(fg)^\wedge = \hat{f} * \hat{g}\ .$$

PROOF. We know from the L_1 theory that

$$(f * g)^\wedge = (2\pi)^{d/2}\hat{f}\hat{g}\ ,$$

so

$$(\hat{f} * \hat{g})^\wedge = (2\pi)^{d/2}\hat{\hat{f}}\hat{\hat{g}} = (2\pi)^{d/2}(fg)^{\wedge\wedge}\ ,$$

since $\mathcal{F}^2$ is reflection. The Fourier inverse then gives

$$\hat{f} * \hat{g} = (2\pi)^{d/2}(fg)^\wedge\ .$$

We saw in Proposition 5.13 that $\check{f}\check{g} \in \mathcal{S}$, so also

$$f * g = \hat{\check{f}} * \hat{\check{g}} = (2\pi)^{d/2}(\check{f}\check{g})^\wedge \in \mathcal{S}\ . \quad \square$$

5.3. The $L_2(\mathbb{R}^d)$ theory

Recall from Proposition 5.6 that for $f, g \in \mathcal{S}$,

$$\int f\hat{g} = \int \hat{f}g\ .$$

COROLLARY 5.19. *If $f, g \in \mathcal{S}$,*

$$\int f(x)\overline{g(x)}\,dx = \int \hat{f}(\xi)\overline{\hat{g}(\xi)}\,d\xi\ .$$

PROOF. We compute

$$\int f\bar{g} = \int f\hat{\check{\bar{g}}} = \int \hat{f}\check{\bar{g}} = \int \hat{f}\bar{\hat{g}}\ ,$$

since $\check{\bar{g}} = \bar{\hat{g}}$ is readily verified. $\square$

Thus $\mathcal{F}$ preserves the L_2 inner product on $\mathcal{S}$. Since $\mathcal{S} \subset L_2(\mathbb{R}^d)$ is dense, we *extend* $\mathcal{F} : \mathcal{S}$ (with L_2 topology) $\to L_2$ to $\mathcal{F} : L_2 \to L_2$ by the following general result.

THEOREM 5.20. *Suppose X and Y are complete metric spaces and $A \subset X$ is dense. If $T : A \to Y$ is uniformly continuous, then there is a unique extension $\tilde{T} : X \to Y$ which is continuous.*

PROOF. Given $x \in X$, take $\{x_j\}_{j=1}^\infty \subset A$ such that $x_j \xrightarrow{X} x$. Let $y_j = T(x_j)$. Since T is uniformly continuous, $\{y_j\}_{j=1}^\infty$ is Cauchy in Y. Let $y_j \xrightarrow{Y} y$ and *define* $\tilde{T}(x) = y = \lim_{j\to\infty} T(x_j)$.

Note that $\tilde{T}$ is well defined since A is dense and limits exist uniquely in a complete metric space. If $\tilde{T}$ is fully continuous (i.e., not just for limits from A), then any other continuous extension would necessarily agree with $\tilde{T}$, so $\tilde{T}$ would be unique.

To see that indeed $\tilde{T}$ is continuous, let $\varepsilon > 0$ be given. Since T is uniformly continuous, there is $\delta > 0$ such that for all $x, \xi \in A$,

$$d_Y(T(x), T(\xi)) < \varepsilon \quad \text{whenever} \quad d_X(x, \xi) < \delta \ .$$

Now let $x, \xi \in X$ such that $d_X(x, \xi) < \delta/3$. Choose $\{x_j\}_{j=1}^\infty$ and $\{\xi_j\}_{j=1}^\infty$ in A such that $x_j \xrightarrow{X} x$ and $\xi_j \xrightarrow{X} \xi$, and choose N large enough so that for $j \geq N$,

$$d_X(x_j, \xi_j) \leq d(x_j, x) + d(x, \xi) + d(\xi, \xi_j) < \delta \ .$$

Then

$$d_Y(\tilde{T}(x), \tilde{T}(\xi)) \leq d_Y(\tilde{T}(x), T(x_j)) + d_Y(T(x_j), T(\xi_j)) + d_Y(T(\xi_j), \tilde{T}(\xi)) < 3\varepsilon \ ,$$

provided j is sufficiently large. That is, $\tilde{T}$ is continuous (but *not* necessarily uniformly so!). $\square$

COROLLARY 5.21. *If X and Y are Banach spaces, $A \subset X$ is dense, and $T : A \to Y$ is continuous and linear, then there is a unique continuous linear extension $\tilde{T} : X \to Y$.*

PROOF. A continuous linear map is uniformly continuous, and the extension, defined by continuity, is necessarily linear. $\square$

THEOREM 5.22 (Plancherel). *The Fourier transform extends to a unitary isomorphism of $L_2(\mathbb{R}^d)$ to itself. That is,*

$$\mathcal{F} : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$$

is a linear, one-to-one, and onto map such that the $L_2(\mathbb{R}^d)$ inner product is preserved:

$$\int f(x)\overline{g(x)}\, dx = \int \hat{f}(\xi)\overline{\hat{g}(\xi)}\, d\xi \ . \tag{5.7}$$

Moreover, $\mathcal{F}^*\mathcal{F} = I$, $\mathcal{F}^* = \mathcal{F}^{-1}$, $\|\mathcal{F}\| = 1$,

$$\|f\|_{L_2} = \|\hat{f}\|_{L_2} \quad \forall\ f \in L_2(\mathbb{R}^d) \ ,$$

and $\mathcal{F}^2$ is reflection.

PROOF. Note that $\mathcal{S}$ (in fact C_0^∞) is dense in $L_2(\mathbb{R}^d)$, and that Corollary 5.19 (i.e., (5.7) on $\mathcal{S}$) implies uniform continuity of $\mathcal{F}$ on $\mathcal{S}$:

$$\|\hat{f}\|_{L_2(\mathbb{R}^d)} = \left(\int \hat{f}\bar{\hat{f}}\, dx)\right)^{1/2} = \left(\int f\bar{f}\, dx\right)^{1/2} = \|f\|_{L_2(\mathbb{R}^d)} \ .$$

We therefore extend $\mathcal{F}$ uniquely to $L_2(\mathbb{R}^d)$ as a continuous operator. Trivially $\mathcal{F}$ is linear and $\|\mathcal{F}\| = 1$. By continuity, (5.7) on $\mathcal{S}$ continues to hold on all of $L_2(\mathbb{R}^d)$ and $\mathcal{F}^*\mathcal{F} = I$.

Similarly we extend $\mathcal{F}^{-1} : \mathcal{S} \to L_2$ to L_2. For $f \in L_2$, $f_j \in \mathcal{S}$, $f_j \to f$ in L_2, we have

$$\mathcal{F}\mathcal{F}^{-1} f = \lim_{j\to\infty} \mathcal{F}\mathcal{F}^{-1} f_j = \lim_{j\to\infty} f_j = f$$

and similarly $\mathcal{F}^{-1}\mathcal{F} f = f$. Thus $\mathcal{F}$ is one-to-one and onto. Since $\mathcal{F}^2$ is reflection on $\mathcal{S}$, it is so on $L_2(\mathbb{R}^d)$ by continuity (or by the uniqueness of the extension, since reflection on $\mathcal{S}$ extends to reflection on $L_2(\mathbb{R}^d)$). $\square$

By the density of $\mathcal{S}$ in $L_2(\mathbb{R}^d)$ and the definition of $\mathcal{F}$ as the continuous extension from $\mathcal{S}$ to L_2, many nice properties of $\mathcal{F}$ on $\mathcal{S}$ extend to $L_2(\mathbb{R}^d)$ trivially.

Corollary 5.23. *For all $f, g \in L_2(\mathbb{R}^d)$,*

$$\int f\hat{g}\,dx = \int \hat{f} g\,dx \ .$$

Proof. Extend Proposition 5.6. $\square$

The following lemma allows us to compute Fourier transforms of L_2 functions.

Lemma 5.24. *Let $f \in L_2(\mathbb{R}^d)$.*

(a) *If $f \in L_1(\mathbb{R}^d)$ as well, then the L_2 Fourier transform of f is*

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

(i.e., the L_1 and L_2 Fourier transforms agree).

(b) *If $R > 0$ and*

$$\varphi_R(\xi) = (2\pi)^{-d/2} \int_{|x|\le R} f(x) e^{-ix\cdot\xi}\,dx \ ,$$

then $\varphi_R \xrightarrow{L_2} \hat{f}$.

Similar statements hold for $\mathcal{F}^{-1}$.

Proof. (a) Since $\mathcal{S} \subset L_1 \cap L_2 \subset L_2$ is dense, we can extend the L_1 Fourier transform from $L_1 \cap L_2$ to L_2. By the uniqueness of the extension, it agrees with the extension from $\mathcal{S}$.

(b) Let $\chi_R(x)$ denote the characteristic function of $B_R(0)$. Then

$$\|\widehat{\chi_R f} - \hat{f}\|_{L_2} = \|\chi_R f - f\|_{L_2} \to 0 \ \text{ as } \ R \to \infty \ . \quad \square$$

5.4. The $\mathcal{S}'$ Theory

The Fourier transform cannot be defined on all distributions, but it can be defined on a subset $\mathcal{S}'$ of $\mathcal{D}'$. Here, $\mathcal{S}'$ is the dual of $\mathcal{S}$. Before attempting the definition, we study $\mathcal{S}$ and $\mathcal{S}'$.

Proposition 5.25. *The inclusion map $i : \mathcal{D} \to \mathcal{S}$ is continuous (i.e., $\mathcal{D} \hookrightarrow \mathcal{S}$, $\mathcal{D}$ is continuously imbedded in $\mathcal{S}$), and $\mathcal{D}$ is dense in $\mathcal{S}$.*

Proof. Suppose that $\phi_j \in \mathcal{D}$ and $\phi_j \to \phi$ in $\mathcal{D}$. Then there is a compact set K such that the supports of the ϕ_j and ϕ are in K, and $\|D^\alpha(\phi_j - \phi)\|_{L_\infty} \to 0$ for every multi-index α. But this immediately implies that in $\mathcal{S}$,

$$\begin{aligned} \rho_n(i(\phi_j) - i(\phi)) &= \sup_{|\alpha|\le n} \sup_{x\in K} (1+|x|^2)^{n/2} |D^\alpha(\phi_j(x) - \phi(x))| \\ &\le \big(\sup_{x\in K} (1+|x|^2)^{n/2}\big) \sup_{|\alpha|\le n} \|D^\alpha(\phi_j - \phi)\|_{L_\infty} \to 0 \ , \end{aligned}$$

since K is bounded, which shows that $i(\phi_j) \to i(\phi)$ in $\mathcal{S}$, i.e., i is continuous.

Let $f \in \mathcal{S}$ and $\phi \in \mathcal{D}$ be such that $\phi \equiv 1$ on $B_1(0)$. For $\varepsilon > 0$, set

$$f_\varepsilon(x) = \phi(\varepsilon x) f(x) \in \mathcal{D} \ .$$

We claim that $f_\varepsilon \xrightarrow{\mathcal{S}} f$, so that $\mathcal{D}$ is dense in $\mathcal{S}$. We need to show that for any multi-indices α and β,

$$\|x^\alpha D^\beta (f - f_\varepsilon)\|_{L_\infty} \to 0 \quad \text{as} \quad \varepsilon \to 0 \ .$$

Now $f(x) = f_\varepsilon(x)$ for $|x| < 1/\varepsilon$, so consider $|x| \geq 1/\varepsilon$. By Leibniz Rule,

$$\begin{aligned}
|x^\alpha D^\beta (f - f_\varepsilon)| &= \Big| x^\alpha \sum_{\gamma \leq \beta} \binom{\beta}{\gamma} D^{\beta-\gamma} f \, D^\gamma (1 - \phi(\varepsilon x)) \Big| \\
&\leq \sum_{\gamma \leq \beta} \binom{\beta}{\gamma} \|x^{\alpha+\delta} D^{\beta-\gamma} f\|_{L_\infty} \|D^\gamma (1 - \phi(\varepsilon x))\|_{L_\infty} \varepsilon^{|\delta|}
\end{aligned}$$

for any multi-index δ. This is uniformly small, so the result follows. $\square$

Corollary 5.26. *If $\phi_j \xrightarrow{\mathcal{D}} \phi$, then $\phi_j \xrightarrow{\mathcal{S}} \phi$.*

Proof. That is, $i(\phi_j) \xrightarrow{\mathcal{S}} i(\phi)$. $\square$

Definition. The dual of $\mathcal{S}$, the space of continuous linear functionals on $\mathcal{S}$, is denoted $\mathcal{S}'$ and called the space of *tempered distributions*.

Proposition 5.27. *Every tempered distribution $u \in \mathcal{S}'$ can be identified naturally with a unique distribution $v \in \mathcal{D}'$ by the relation*

$$v = u \circ i = u|_{\mathcal{D}} \ ;$$

that is, the dual operator $i' : \mathcal{S}' \hookrightarrow \mathcal{D}'$ is the restriction operator, restricting the domain from $\mathcal{S}$ to $\mathcal{D}$, and i' is a one-to-one map.

Proof. If we define $v = u \circ i$, then $v \in \mathcal{D}'$, since i is continuous and linear. If $u, w \in \mathcal{S}'$ and $u \circ i = w \circ i$, then in fact $u = w$ since $\mathcal{D}$ is dense in $\mathcal{S}$. $\square$

Corollary 5.28. *The dual space $\mathcal{S}'$ is precisely the subspace of $\mathcal{D}'$ consisting of those functionals that have continuous extensions from $\mathcal{D}$ to $\mathcal{S}$. Moreover, these extensions are unique.*

Example. If α is any multi-index, then

$$D^\alpha \delta_0 \in \mathcal{S}' \ .$$

We can see easily that $D^\alpha \delta_0$ is continuous as follows. Let $\psi \in \mathcal{D}$ be identically one on a neighborhood of 0. Then for $\phi \in \mathcal{S}$,

$$D^\alpha \delta_0(\psi\phi) = (-1)^{|\alpha|} D^\alpha \phi(0)$$

is well defined, so $D^\alpha \delta_0 : \mathcal{S} \to \mathbb{F}$ is the composition of multiplication by ψ (taking $\mathcal{S}$ to $\mathcal{D}$) and $D^\alpha \delta_0 : \mathcal{D} \to \mathbb{F}$. The latter is continuous. For the former, if $\phi_j \xrightarrow{\mathcal{S}} \phi$, then each $\psi\phi_j$ is supported in $\operatorname{supp}(\psi)$ and $D^\alpha(\psi\phi_j) \xrightarrow{L_\infty} D^\alpha(\psi\phi)$ for all α. Thus $\psi\phi_j \xrightarrow{\mathcal{D}} \psi\phi$, so multiplication by ψ is a continuous operation.

We have the following characterization of $\mathcal{S}'$.

THEOREM 5.29. *Let u be a linear functional on $\mathcal{S}$. Then $u \in \mathcal{S}'$ if and only if there are $C > 0$ and $N \geq 0$ such that*

$$|u(\phi)| \leq C\rho_N(\phi) \quad \forall\ \phi \in \mathcal{S}\ ,$$

where (5.5) *defines $\rho_N(\phi)$.*

PROOF. By linearity, u is continuous if and only if it is continuous at 0. If $\phi_j \in \mathcal{S}$ converges to 0 and we assume the existence of $C > 0$ and $N \geq 0$ such that

$$|u(\phi_j)| \leq C\rho_N(\phi_j) \to 0\ ,$$

we see that u is continuous.

Conversely, suppose that no such $C > 0$ and $N \geq 0$ exist. Then for each $j > 0$, we can find $\psi_j \in \mathcal{S}$ such that $\rho_j(\psi_j) = 1$ and

$$|u(\psi_j)| \geq j\ .$$

Let $\phi_j = \psi_j/j$, so that $\phi_j \to 0$ in $\mathcal{S}$ (since the ρ_n are nested, the tail of the sequence $\rho_n(\phi_j) \leq \rho_j(\phi_j) \leq 1/j$ is eventually small for large j and any fixed n). But u continuous implies that $|u(\phi_j)| \to 0$, which contradicts the previous fact that $|u(\phi_j)| = |u(\psi_j)|/j \geq 1$. ☐

EXAMPLE (Tempered L_p). If for some $N > 0$ and $1 \leq p \leq \infty$,

$$\frac{f(x)}{(1+|x|^2)^{N/2}} \in L_p(\mathbb{R}^d)\ ,$$

then we say that $f(x)$ is a *tempered L_p function* (if $p = \infty$, we also say that f is *slowly increasing*). Define $\Lambda_f \in \mathcal{S}'$ by

$$\Lambda_f(\phi) = \int f(x)\phi(x)\,dx\ .$$

This is well defined since by Hölder's inequality for $1/p + 1/q = 1$,

$$\begin{aligned}
|\Lambda_f(\phi)| &= \Big|\int \frac{f(x)}{(1+|x|^2)^{N/2}}(1+|x|^2)^{N/2}\phi(x)\,dx\Big| \\
&\leq \Big\|\frac{f(x)}{(1+|x|^2)^{N/2}}\Big\|_{L_p}\|(1+|x|^2)^{N/2}\phi(x)\|_{L_q}
\end{aligned}$$

is finite if $q = \infty$ (i.e., $p = 1$), and for $q < \infty$,

$$\begin{aligned}
\|(1+|x|^2)^{N/2}\phi\|_{L_q}^q &= \int (1+|x|^2)^{Nq/2}|\phi(x)|^q\,dx \\
&= \int (1+|x|^2)^{Nq/2-M}(1+|x|^2)^M|\phi(x)|^q\,dx \\
&\leq \Big(\int (1+|x|^2)^{Nq/2-M}\,dx\Big)\|(1+|x|^2)^{M/q}\phi\|_{L_\infty}^q \\
&\leq (C\rho_{M/q}(\phi))^q
\end{aligned}$$

is finite provided M is large enough. By the previous theorem, Λ_f is also continuous, so indeed $\Lambda_f \in \mathcal{S}'$. Since each of the following spaces is in tempered L_p for some p, we have shown:

(a) $L_p(\mathbb{R}^d) \subset \mathcal{S}'$ for all $1 \leq p \leq \infty$;
(b) $\mathcal{S} \subset \mathcal{S}'$;
(c) a polynomial, and more generally any measurable function majorized by a polynomial, is a tempered distribution.

EXAMPLE. Not every function in $L_{1,\mathrm{loc}}(\mathbb{R}^d)$ is in $\mathcal{S}'$. The reader can readily verify that $e^x \notin \mathcal{S}'$ by considering $\phi \in \mathcal{S}$ such that the tail looks like $e^{-|x|/2}$.

Generally we endow $\mathcal{S}'$ with the weak-$*$ topology, so that

$$u_j \xrightarrow{\mathcal{S}'} u \ \text{ if and only if } \ u_j(\phi) \to u(\phi) \quad \forall\, \phi \in \mathcal{S}\ .$$

PROPOSITION 5.30. *For any $1 \le p \le \infty$, $L_p \hookrightarrow \mathcal{S}'$ (L_p is continuously imbedded in $\mathcal{S}'$).*

PROOF. We need to show that if $f_j \xrightarrow{L_p} f$, then

$$\int (f_j - f)\phi\, dx \to 0 \quad \forall\, \phi \in \mathcal{S}\ ,$$

which is true by Hölder's inequality. $\square$

As with distributions, we can define operations on tempered distributions by duality: if $T : \mathcal{S} \to \mathcal{S}$ is continuous, and linear, then so is $T' : \mathcal{S}' \to \mathcal{S}'$. Since $\mathcal{F} : \mathcal{S} \to \mathcal{S}$ is continuous linear, we define the Fourier transform on $\mathcal{S}'$ this way.

PROPOSITION 5.31. *If α is a multi-index, $x \in \mathbb{R}^d$, and $f \in C^\infty(\mathbb{R}^d)$ is such that $D^\beta f$ grows at most polynomially for all β, then for $u \in \mathcal{S}'$ and all $\phi \in \mathcal{S}$, the following hold.*

(a) $\langle D^\alpha u, \phi\rangle \equiv \langle u, (-1)^{|\alpha|} D^\alpha \phi\rangle$ *defines* $D^\alpha u \in \mathcal{S}'$.
(b) $\langle fu, \phi\rangle \equiv \langle u, f\phi\rangle$ *defines* $fu \in \mathcal{S}'$.
(c) $\langle \tau_x u, \phi\rangle \equiv \langle u, \tau_{-x}\phi\rangle$ *defines* $\tau_x u \in \mathcal{S}'$.
(d) $\langle Ru, \phi\rangle \equiv \langle u, R\phi\rangle$, *where R is reflection about $x = 0$, defines* $Ru \in \mathcal{S}'$.
(e) $\langle \hat{u}, \phi\rangle \equiv \langle u, \hat{\phi}\rangle$ *defines* $\hat{u} \in \mathcal{S}'$.
(f) $\langle \check{u}, \phi\rangle \equiv \langle u, \check{\phi}\rangle$ *defines* $\check{u} \in \mathcal{S}'$.

Moreover, these operations are continuous on $\mathcal{S}'$.

Note that if $\phi \in \mathcal{D}$, then $\hat{\phi} \notin \mathcal{D}$, so the Fourier transform $\mathcal{F}$ is *not* defined for all $u \in \mathcal{D}'$. We also have convolution defined for $u \in \mathcal{S}'$ and $\phi \in \mathcal{S}$:

$$(u * \phi)(x) = \langle u, \tau_x R\phi\rangle\ .$$

PROPOSITION 5.32. *For $u \in \mathcal{S}'$ and $\phi \in \mathcal{S}$,*

(a) $u * \phi \in C^\infty$ *and*

$$D^\alpha(u * \phi) = (D^\alpha u) * \phi = u * D^\alpha \phi \quad \forall\ \alpha\ ,$$

(b) $u * \phi \in \mathcal{S}'$ *(in fact, $u * \phi$ grows at most polynomially).*

PROOF. The proof of (a) is similar to the case of distributions and left to the reader. For (b), note that

$$1 + |x + y|^2 \le 2(1 + |x|^2)(1 + |y|^2)\ ,$$

so

$$\rho_N(\tau_x \phi) \le 2^{N/2}(1 + |x|^2)^{N/2} \rho_N(\phi)\ .$$

Now $u \in \mathcal{S}'$, so there are $C > 0$ and $N \ge 0$ such that

$$|u(\phi)| \le C\rho_N(\phi)\ ,$$

so

$$|u * \phi| = |u(\tau_x R\phi)| \le C 2^{N/2}(1 + |x|^2)^{N/2} \rho_N(\phi)$$

shows $u * \phi \in \mathcal{S}'$ and grows at most polynomially. $\square$

Let us study the Fourier transform of tempered distributions. Recall that if f is a tempered L_p function, then $\Lambda_f \in \mathcal{S}'$.

PROPOSITION 5.33. *If $f \in L_1 \cup L_2$, then $\hat{\Lambda}_f = \Lambda_{\hat{f}}$ and $\check{\Lambda}_f = \Lambda_{\check{f}}$. That is, the L_1 and L_2 definitions of the Fourier transform are consistent with the $\mathcal{S}'$ definition.*

PROOF. For $\phi \in \mathcal{S}$,

$$\langle \hat{\Lambda}_f, \phi\rangle = \langle \Lambda_f, \hat{\phi}\rangle = \int f\hat{\phi} = \int \hat{f}\phi = \langle \Lambda_{\hat{f}}, \phi\rangle \ ,$$

so $\hat{\Lambda}_f = \Lambda_{\hat{f}}$. A similar computation gives the result for the Fourier inverse transform. □

PROPOSITION 5.34. *If $u \in \mathcal{S}'$, then*
(a) $\check{\hat{u}} = u$,
(b) $\hat{\check{u}} = u$,
(c) $\hat{\hat{u}} = Ru$,
(d) $\hat{u} = (Ru)^\vee = R\check{u}$.

PROOF. By definition, since these hold on $\mathcal{S}$. □

THEOREM 5.35 (Plancherel). *The Fourier transform is a continuous, linear, one-to-one mapping of $\mathcal{S}'$ onto $\mathcal{S}'$, of period 4, with a continuous inverse.*

PROOF. If $u_j \xrightarrow{\mathcal{S}'} u$, (i.e., $\langle u_j, \phi\rangle \to \langle u, \phi\rangle$ for all $\phi \in \mathcal{S}$), then

$$\langle \hat{u}_j, \phi\rangle = \langle u_j, \hat{\phi}\rangle \to \langle u, \hat{\phi}\rangle = \langle \hat{u}, \phi\rangle \ ,$$

so $\hat{u}_j \to \hat{u}$; that is, the Fourier transform is continuous. Now

$$\mathcal{F}^2 u = \hat{\hat{u}} = Ru \ ,$$

so

$$\mathcal{F}^4 u = R^2 u = u = \mathcal{F}(\mathcal{F}^3)u = (\mathcal{F}^3)\mathcal{F}u$$

shows that $\mathcal{F}$ has period 4 and has a continuous inverse $\mathcal{F}^{-1} = \mathcal{F}^3$. □

EXAMPLE. Consider $\delta_0 \in \mathcal{S}'$. For $\phi \in \mathcal{S}$,

$$\langle \hat{\delta}_0, \phi\rangle = \langle \delta_0, \hat{\phi}\rangle = \hat{\phi}(0) = (2\pi)^{-d/2}\int \phi(x)\,dx = \langle (2\pi)^{-d/2}, \phi\rangle \ ,$$

so

$$\hat{\delta}_0 = (2\pi)^{-d/2} \ .$$

Conversely, by Proposition 5.34(d),

$$\delta_0 = \mathcal{F}^{-1}(2\pi)^{-d/2} = \mathcal{F}(2\pi)^{-d/2} \ ,$$

so

$$\hat{1} = (2\pi)^{d/2}\delta_0 \ .$$

PROPOSITION 5.36. *If $u \in \mathcal{S}'$, $y \in \mathbb{R}^d$, and α is a multi-index, then*
(a) $(\tau_y u)^\wedge = e^{-iy\cdot\xi}\hat{u}$,
(b) $\tau_y \hat{u} = (e^{iy\cdot x}u)^\wedge$,
(c) $(D^\alpha u)^\wedge = (i\xi)^\alpha \hat{u}$,

(d) $D^\alpha \hat{u} = \big((-i\xi)^\alpha u\big)^\wedge$.

Proposition 5.31 (b) implies that the products involving tempered distributions are well defined in $\mathcal{S}'$.

PROOF. For (a), consider $\phi \in \mathcal{S}$ and

$$\langle (\tau_y u)^\wedge, \phi\rangle = \langle \tau_y u, \hat{\phi}\rangle = \langle u, \tau_{-y}\hat{\phi}\rangle = \langle u, \widehat{e^{-iy\cdot\xi}\phi}\rangle = \langle \hat{u}, e^{-iy\cdot\xi}\phi\rangle = \langle e^{-iy\cdot\xi}\hat{u}, \phi\rangle \ .$$

Results (b)–(d) are shown similarly. □

PROPOSITION 5.37. *If $u \in \mathcal{S}'$ and $\phi, \psi \in \mathcal{S}$, then*
(a) $(u * \phi)^\wedge = (2\pi)^{d/2}\hat{\phi}\hat{u}$,
(b) $(u * \phi) * \psi = u * (\phi * \psi)$.

PROOF. Let $\check{\psi} \in \mathcal{S}$ and choose $\psi_j \in \mathcal{D}$ with support in K_j such that $\psi_j \xrightarrow{\mathcal{S}} \psi$ (so also $\check{\psi}_j \xrightarrow{\mathcal{S}} \check{\psi}$). Now

$$\langle (u*\phi)^\wedge, \check{\psi}_j\rangle = \langle u*\phi, \psi_j\rangle = \int u*\phi(x)\psi_j(x)\,dx \ ,$$

since $u * \phi \in C^\infty$ and has polynomial growth. Continuing, this is

$$\int_{K_j} \langle u, \tau_x R\phi\rangle \psi_j(x)\,dx = \left\langle u, \int_{K_j} \tau_x R\phi\psi_j(x)\,dx \right\rangle \ ,$$

which we see by approximating the integral by Riemann sums and using the linearity and continuity of u. Continuing, this is

$$\begin{aligned}
\left\langle u, \int \phi(x-y)\psi_j(x)\,dx \right\rangle &= \langle u, R\phi * \psi_j\rangle \\
&= \langle \hat{u}, (R\phi * \psi_j)^\vee\rangle \\
&= (2\pi)^{d/2}\langle \hat{u}, (R\phi)^\vee \check{\psi}_j\rangle \\
&= (2\pi)^{d/2}\langle \hat{\phi}\hat{u}, \check{\psi}_j\rangle \\
&\to (2\pi)^{d/2}\langle \hat{\phi}\hat{u}, \check{\psi}\rangle \ .
\end{aligned}$$

That is, for all $\check{\psi} \in \mathcal{S}$,

$$\langle (u*\phi)^\wedge, \check{\psi}\rangle = \langle (2\pi)^{d/2}\hat{\phi}\hat{u}, \check{\psi}\rangle \ ,$$

and (a) follows.

Finally, (b) follows from (a):

$$((u*\phi)*\psi)^\wedge = (2\pi)^{d/2}\hat{\psi}(u*\phi)^\wedge = (2\pi)^d\hat{\psi}\hat{\phi}\hat{u}$$

and

$$(u*(\phi*\psi))^\wedge = (2\pi)^{d/2}(\phi*\psi)^\wedge\hat{u} = (2\pi)^d\hat{\phi}\hat{\psi}\hat{u} \ .$$

Thus

$$((u*\phi)*\psi)^\wedge = (u*(\phi*\psi))^\wedge \ ,$$

and the Fourier inverse gives (b). □

Example (Heat operator). The heat operator for $(x,t) \in \mathbb{R}^d \times (0,\infty)$ is

$$\frac{\partial}{\partial t} - \Delta \ .$$

It models the flow of heat in space and time. We consider the initial value problem (IVP)

$$\begin{cases} \dfrac{\partial u}{\partial t} - \Delta u = 0 \ , & (x,t) \in \mathbb{R}^d \times (0,\infty), \\ u(x,0) = f(x), & x \in \mathbb{R}^d, \end{cases}$$

where $f(x)$ is given. To find a solution, we proceed *formally* (i.e., without rigor). Assume that the solution is at least a tempered distribution and take the Fourier transform in x only, for each fixed t:

$$\begin{cases} \widehat{\dfrac{\partial u}{\partial t}} - \widehat{\Delta u} = \dfrac{\partial}{\partial t}\hat{u} + |\xi|^2 \hat{u} = 0 \ , \\ \hat{u}(\xi,0) = \hat{f}(\xi) \ . \end{cases}$$

For each fixed $\xi \in \mathbb{R}^d$, this is an ordinary differential equation with an initial condition. Its solution is

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

Thus, using Lemma 5.16 and Proposition 5.37,

$$\begin{aligned} u(x,t) &= (\hat{f} e^{-|\xi|^2 t})^\vee \\ &= \left[\hat{f} \Big(\frac{1}{(2t)^{d/2}} e^{-|x|^2/4t} \Big)^\wedge \right]^\vee \\ &= (2\pi)^{-d/2} f * \Big(\frac{1}{(2t)^{d/2}} e^{-|x|^2/4t} \Big) \ . \end{aligned}$$

Define the *Gaussian*, or *heat*, *kernel*

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

Then

$$u(x,t) = (f * K(\cdot,t))(x)$$

should be a solution to our IVP, and K should solve the IVP with $f = \delta_0$. In fact,

$$\int K(x,t)\,dx = \hat{K}(0,t) = 1 \ \forall \ t$$

and

$$K(x,t) = t^{-d/2} K(t^{-1/2}x, 1) \ ,$$

so K approximates δ_0 as $t \to 0$. Thus the initial condition is satisfied as $t \to 0^+$, and K controls how the initial condition (initial heat distribution) dissipates with time. To remove the formality of the above calculation, we start with $K(x,t)$ defined as above, and note that for $f \in \mathcal{D}$ and $u = f * K$ as above,

$$u_t - \Delta u = f * (K_t - \Delta K) = f * 0 = 0 \ .$$

To extend to $f \in L_p$, we use that $\mathcal{D}$ is dense in L_p. See [**Fo**, p. 190] for details.

5.5. Exercises

1. Compute the Fourier transform of $e^{-|x|}$ for $x \in \mathbb{R}$.

2. Compute the Fourier transform of $e^{-a|x|^2}$, $a > 0$, directly, where $x \in \mathbb{R}$. You will need to use the Cauchy Theorem.

3. If $f \in L_1(\mathbb{R}^d)$ and $f > 0$, show that for every $\xi \neq 0$, $|\hat{f}(\xi)| < \hat{f}(0)$.

4. If $f \in L_1(\mathbb{R}^d)$ and $f(x) = g(|x|)$ for some g, show that $\hat{f}(\xi) = h(|\xi|)$ for some h. Can you relate g and h?

5. Give an example of a function $f \in L_2(\mathbb{R}^d)$ which is not in $L_1(\mathbb{R}^d)$, but such that $\hat{f} \in L_1(\mathbb{R}^d)$. Under what circumstances can this happen?

6. Suppose that $f \in L_p(\mathbb{R}^d)$ for some p between 1 and 2.

   (a) Show that there are $f_1 \in L_1(\mathbb{R}^d)$ and $f_2 \in L_2(\mathbb{R}^d)$ such that $f = f_1 + f_2$.

   (b) Define $\hat{f} = \hat{f}_1 + \hat{f}_2$. Show that this definition is well defined; that is, that it is independent of the choice of f_1 and f_2.

7. Suppose that f and g are in $L_2(\mathbb{R}^d)$. The convolution $f * g$ is in $L_\infty(\mathbb{R}^d)$, so it may not have a Fourier transform. Nevertheless, prove that $f * g = (2\pi)^{d/2}(\hat{f}\hat{g})^\vee$ is well defined, wherein the Fourier inverse is given by the usual integration formula.

8. Find the eigenvalues of the Fourier transform: $\hat{f} = \lambda f$.

9. Compute the Fourier Transforms of the following functions, considered as tempered distributions.

   (a) $f(x) = x$ for $x \in \mathbb{R}$.

   (b) $g(x) = e^{-|x|}$ for $x \in \mathbb{R}$.

   (c) $h(x) = e^{i|x|^2}$ for $x \in \mathbb{R}^d$.

   (d) $\sin x$ and $\cos x$ for $x \in \mathbb{R}$.

10. Let $\varphi \in \mathcal{S}(\mathbb{R}^d)$, $\hat{\varphi}(0) = (2\pi)^{-d/2}$, and $\varphi_\epsilon(x) = \epsilon^{-n}\varphi(x/\epsilon)$. Prove that $\varphi_\epsilon \to \delta_0$ and $\hat{\varphi}_\epsilon \to (2\pi)^{-d/2}$ as $\epsilon \to 0^+$. In what sense do these convergences take place?

11. Let $1 \le p < \infty$ and suppose $f \in L_p(\mathbb{R})$. Let $g(x) = \displaystyle\int_x^{x+1} f(y)\,dy$. Prove that $g \in C_v(\mathbb{R})$.

12. Show that the Fourier Transform $\mathcal{F} : L_1(\mathbb{R}^d) \to C_v(\mathbb{R}^d)$ is not onto. Show, however, that $\mathcal{F}(L_1(\mathbb{R}^d))$ is dense in $C_v(\mathbb{R}^d)$. [Hint: See Exercise 11.]

13. Is it possible for there to be a continuous function f defined on $\mathbb{R}^d$ with the following two properties?

    (a) There is no polynomial P in d variables such that $|f(x)| \le P(x)$ for all $x \in \mathbb{R}^d$.

    (b) The distribution $\phi \mapsto \int \phi\, f\, dx$ is tempered.

14. Let the field be complex and define $T : L_2(\mathbb{R}^d) \to L_2(\mathbb{R}^d)$ by

$$Tf(x) = \int e^{-|x-y|^2/2} f(y)\, dy\ .$$

Use the Fourier transform to show that T is a positive, injective operator, but that T is *not* surjective.

15. When is $\sum_{k=1}^{\infty} a_k\delta_k \in \mathcal{S}'(\mathbb{R})$? (Here, δ_k is the point mass centered at $x = k$.)

16. For $f \in L_2(\mathbb{R})$, define the Hilbert transform of f by $Hf = \mathrm{PV}\left(\frac{1}{\pi x}\right) * f$, where the convolution uses ordinary Lebesgue measure.
   (a) Show that $\mathcal{F}(\mathrm{PV}(1/x)) = -i\sqrt{\pi/2}\,\mathrm{sgn}(\xi)$, where $\mathrm{sgn}(\xi)$ is the sign of ξ.
   (b) Show that $\|Hf\|_{L_2} = \|f\|_{L_2}$ and $HHf = -f$.

17. Let T be a bounded linear transformation mapping $L_2(\mathbb{R}^d)$ into itself. If there exists a bounded measurable function $m(\xi)$ (a *multiplier*) such that $\widehat{Tf}(\xi) = m(\xi)\hat{f}(\xi)$ for all $f \in L_2(\mathbb{R}^d)$, show that then T commutes with translation and $\|T\| = \|m\|_{L_\infty}$. Such operators are called multiplier operators. (Remark: the converse of this statement is also true.)

18. Give a careful argument that $\mathcal{D}(\mathbb{R}^d)$ is dense in $\mathcal{S}$. Show also that $\mathcal{S}'$ is dense in $\mathcal{D}'$ and that distributions with compact support are dense in $\mathcal{S}'$.

19. Make an argument that there is no simple way to define the Fourier transform on $\mathcal{D}'$ in the way we have for $\mathcal{S}'$.

20. Use the Fourier Transform to find a solution to
$$u - \frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2} = e^{-x_1^2 - x_2^2}\ .$$
Hint: write your answer in terms of a suitable inverse Fourier transform and a convolution. Can you find a fundamental solution to the differential operator?

CHAPTER 6

Sobolev Spaces

In this chapter we define and study some important families of Banach spaces of measurable functions with distributional derivatives that lie in some L_p space ($1 \le p \le \infty$). We include spaces of "fractional order" of functions having smoothness between integral numbers of derivatives, as well as their dual spaces, which contain elements that lack derivatives.

While such spaces arise in a number of contexts, one basic motivation for their study is to understand the trace of a function. Consider a domain $\Omega \subset \mathbb{R}^d$ and its boundary $\partial\Omega$. If $f \in C^0(\bar{\Omega})$, then its trace $f|_{\partial\Omega}$ is well defined and $f|_{\partial\Omega} \in C^0(\partial\Omega)$. However, if merely $f \in L_2(\Omega)$, then $f|_{\partial\Omega}$ is *not* defined, since $\partial\Omega$ has measure zero in $\mathbb{R}^d$. That is, f is actually the equivalence class of all functions on Ω that differ on a set of measure zero from any other function in the class; thus, $f|_{\partial\Omega}$ can be chosen arbitrarily from the equivalence class. As part of what we will see, if $f \in L_2(\Omega)$ *and* $\partial f/\partial x_i \in L_2(\Omega)$ for $i = 1, \dots, d$, then in fact $f|_{\partial\Omega}$ can be defined uniquely, and, in fact, $f|_{\partial\Omega}$ has $1/2$ derivative.

6.1. Definitions and Basic Properties

We begin by defining Sobolev spaces of functions with an integral number of derivatives.

Definition (Sobolev Spaces). Let $\Omega \subset \mathbb{R}^d$ be a domain, $1 \le p \le \infty$, and $m \ge 0$ be an integer. The *Sobolev space* of m derivatives in $L_p(\Omega)$ is

$$W^{m,p}(\Omega) = \{f \in L_p(\Omega) : D^\alpha f \in L_p(\Omega) \text{ for all multi-indices } \alpha \text{ such that } |\alpha| \le m\} .$$

Of course, the elements are equivalence classes of functions that differ only on a set of measure zero. The derivatives are taken in the sense of distributions.

Example. The reader can verify that when Ω is bounded, $f(x) = |x|^\alpha \in W^{m,p}(\Omega)$ if and only if $(\alpha - m)p + d > 0$.

Definition. For $f \in W^{m,p}(\Omega)$, the $W^{m,p}(\Omega)$-*norm* is

$$\|f\|_{W^{m,p}(\Omega)} = \left\{ \sum_{|\alpha| \le m} \|D^\alpha f\|^p_{L_p(\Omega)} \right\}^{1/p} \quad \text{if} \quad p < \infty$$

and

$$\|f\|_{W^{m,\infty}(\Omega)} = \max_{|\alpha| \le m} \|D^\alpha f\|_{L_\infty(\Omega)} \quad \text{if} \quad p = \infty .$$

Proposition 6.1.

(a) $\|\cdot\|_{W^{m,p}(\Omega)}$ *is indeed a norm.*
(b) $W^{0,p}(\Omega) = L_p(\Omega)$.
(c) $W^{m,p}(\Omega) \hookrightarrow W^{k,p}(\Omega)$ *for all* $m \ge k \ge 0$ *(i.e.,* $W^{m,p}$ *is* continuously *imbedded in* $W^{k,p}$*).*

The proof is easy and left to the reader.

PROPOSITION 6.2. *The space $W^{m,p}(\Omega)$ is a Banach space.*

PROOF. It remains to show that $W^{m,p}(\Omega)$ is complete. Let $\{u_j\}_{j=1}^{\infty} \subset W^{m,p}(\Omega)$ be Cauchy. Then $\{D^\alpha u_j\}_{j=1}^{\infty}$ is Cauchy in $L_p(\Omega)$ for all $|\alpha| \leq m$, and, $L_p(\Omega)$ being complete, there are functions $u_\alpha \in L_p(\Omega)$ such that

$$D^\alpha u_j \xrightarrow{L_p} u_\alpha \quad \text{as} \quad j \to \infty \ .$$

We let $u = u_0$ and claim that $D^\alpha u = u_\alpha$. To see this, let $\phi \in \mathcal{D}$ and note that

$$\langle D^\alpha u_j, \phi \rangle \to \langle u_\alpha, \phi \rangle$$

and

$$\langle D^\alpha u_j, \phi \rangle = (-1)^{|\alpha|} \langle u_j, D^\alpha \phi \rangle \longrightarrow (-1)^{|\alpha|} \langle u, D^\alpha \phi \rangle = \langle D^\alpha u, \phi \rangle \ .$$

Thus $u_\alpha = D^\alpha u$ as distributions, and so also as $L_p(\Omega)$ functions. We conclude that

$$D^\alpha u_j \xrightarrow{L_p} D^\alpha u \quad \forall\ |\alpha| \leq m \ ;$$

that is,

$$u_j \xrightarrow{W^{m,p}} u \ .$$

□

Certain basic properties of L_p spaces hold for $W^{m,p}$ spaces.

PROPOSITION 6.3. *The space $W^{m,p}(\Omega)$ is separable if $1 \leq p < \infty$ and reflexive if $1 < p < \infty$.*

PROOF. We use strongly the same result known for $L_p(\Omega)$, i.e., $m = 0$. Let N denote the number of multi-indices of order less than or equal to m. Let

$$L_p^N = \underbrace{L_p(\Omega) \times \cdots \times L_p(\Omega)}_{N \text{ times}} = \prod_{j=1}^{N} L_p(\Omega)$$

and define the norm for $u \in L_p^N$ by

$$\|u\|_{L_p^N} = \left\{ \sum_{j=1}^{N} \|u_j\|_{L_p(\Omega)}^p \right\}^{1/p} \ .$$

It is trivial to verify that L_p^N is a Banach space with properties similar to those of L_p: L_p^N is separable and reflexive if $p > 1$, since $(L_p^N)^* = L_q^N$ where $1/p + 1/q = 1$. Define $T : W^{m,p}(\Omega) \to L_p^N$ by

$$(Tu)_j = D^\alpha u \ ,$$

where α is the j^{th} multi-index. Then T is linear and

$$\|Tu\|_{L_p^N} = \|u\|_{W^{m,p}(\Omega)} \ .$$

That is, T is an isometric isomorphism of $W^{m,p}(\Omega)$ onto a subspace W of L_p^N. Since $W^{m,p}(\Omega)$ is complete, W is closed. Thus, since L_p^N is separable, so is W, and since L_p^N is reflexive for $1 < p < \infty$, so is W. □

When $p = 2$, we have a Hilbert space.

DEFINITION. We denote the m^{th} order Sobolev space in $L_2(\Omega)$ by

$$H^m(\Omega) = W^{m,2}(\Omega) \ .$$

PROPOSITION 6.4. *The space $H^m(\Omega) = W^{m,2}(\Omega)$ is a separable Hilbert space with the inner product*

$$(u,v)_{H^m(\Omega)} = \sum_{|\alpha|\le m} (D^\alpha u, D^\alpha v)_{L_2(\Omega)} \ ,$$

where

$$(f,g)_{L_2(\Omega)} = \int_\Omega f(x)\overline{g(x)}\,dx$$

is the usual $L_2(\Omega)$ inner product.

When $p < \infty$, a very useful fact about Sobolev spaces is that C^∞ functions form a dense subset. In fact, one can *define* $W^{m,p}(\Omega)$ to be the *completion* (i.e., the set of "limits" of Cauchy sequences) of $C^\infty(\Omega)$ (or even $C^m(\Omega)$) with respect to the $W^{m,p}(\Omega)$-norm.

THEOREM 6.5. *If $1 \le p < \infty$, then*

$$\{f \in C^\infty(\Omega) : \|f\|_{W^{m,p}(\Omega)} < \infty\} = C^\infty(\Omega) \cap W^{m,p}(\Omega)$$

is dense in $W^{m,p}(\Omega)$.

We need several results before we can prove this theorem.

LEMMA 6.6. *Suppose that $1 \le p < \infty$ and $\varphi \in C_0^\infty(\mathbb{R}^d)$ is an approximate identity supported in the unit ball about the origin (i.e., $\varphi \ge 0$, $\int \varphi(x)\,dx = 1$, $\operatorname{supp}(\varphi) \subset B_1(0)$, and $\varphi_\varepsilon(x) = \varepsilon^{-d}\varphi(\varepsilon^{-1}x)$ for $\varepsilon > 0$). If $f \in L_p(\Omega)$ is extended by 0 to $\mathbb{R}^d$ (if necessary), then*

(a) $\varphi_\varepsilon * f \in L_p(\mathbb{R}^d)$,
(b) $\|\varphi_\varepsilon * f\|_{L_p} \le \|f\|_{L_p}$,
(c) $\varphi_\varepsilon * f \xrightarrow{L_p} f$ *as* $\varepsilon \to 0^+$.

PROOF. Conclusions (a) and (b) follow from Young's inequality. For (c), we use the fact that continuous functions with compact support are dense in $L_p(\mathbb{R}^d)$. Let $\eta > 0$ and choose $g \in C_0(\mathbb{R}^d)$ such that

$$\|f - g\|_{L_p} \le \eta/3 \ .$$

Then, using (b),

$$\begin{aligned}\|\varphi_\varepsilon * f - f\|_{L_p} &\le \|\varphi_\varepsilon * (f-g)\|_{L_p} + \|\varphi_\varepsilon * g - g\|_{L_p} + \|g - f\|_{L_p}\\ &\le 2\eta/3 + \|\varphi_\varepsilon * g - g\|_{L_p} \ .\end{aligned}$$

Since g has compact support, it is uniformly continuous. Now $\operatorname{supp}(g) \subset B_R(0)$, so $\operatorname{supp}(\varphi_\varepsilon * g - g) \subset B_{R+2}(0)$ for all $\varepsilon \le 1$. Choose $0 < \varepsilon \le 1$ such that

$$|g(x) - g(y)| \le \frac{\eta}{3|B_{R+2}(0)|^{1/p}}$$

whenever $|x - y| < 2\varepsilon$, where $|B_{R+2}(0)|$ is the measure of the ball. Then for $x \in B_{R+2}(0)$,

$$\begin{aligned}(\varphi_\varepsilon * g - g)(x) &= \int \varphi_\varepsilon(x-y)(g(y) - g(x))\,dy\\ &\le \sup_{|x-y|<2\varepsilon} |g(y) - g(x)| \le \frac{\eta}{3|B_{R+2}(0)|^{1/p}} \ ,\end{aligned}$$

so $\|\varphi_\varepsilon * g - g\|_{L_p} \le \eta/3$ and $\|\varphi_\varepsilon * f - f\|_{L_p} \le \eta$ is as small as we like. $\square$

COROLLARY 6.7. *If* $\Omega' \subset\subset \Omega$ *or* $\Omega' = \Omega = \mathbb{R}^d$, *then*

$$\varphi_\varepsilon * f \xrightarrow{W^{m,p}(\Omega')} f \quad \forall\ f \in W^{m,p}(\Omega) \ .$$

PROOF. Extend f by 0 to $\mathbb{R}^d$ if necessary. For any multi-index α with $|\alpha| \le m$,

$$D^\alpha(\varphi_\varepsilon * f) = \varphi_\varepsilon * D^\alpha f \ ,$$

since $\varphi_\varepsilon \in \mathcal{D}(\mathbb{R}^d)$ and $f \in \mathcal{D}'(\mathbb{R}^d)$. The subtlety above is whether $D^\alpha f$, on $\mathbb{R}^d$ after extension of f, has a δ-function on $\partial\Omega$; however, restriction to Ω' removes any difficulty:

$$\varphi_\varepsilon * D^\alpha f \xrightarrow{L_p(\Omega')} D^\alpha f \ ,$$

since eventually as $\varepsilon \to 0$, $\varphi_\varepsilon * D^\alpha f$ involves only values of $D^\alpha f$ strictly supported in Ω. $\square$

PROOF OF THEOREM 6.5. Define $\Omega_0 = \Omega_{-1} = \emptyset$ for integer $k \ge 1$

$$\Omega_k = \{x \in \Omega : |x| < k \ \text{ and } \ \operatorname{dist}(x, \partial\Omega) > 1/k\} \ .$$

Let $\phi_k \in C_0^\infty(\Omega)$ be such that $0 \le \phi_k \le 1$, $\phi_k \equiv 1$ on Ω_k, and $\phi_k \equiv 0$ on Ω_{k+1}^c. Let $\psi_1 = \phi_1$ and $\psi_k = \phi_k - \phi_{k-1}$ for $k \ge 2$, so $\psi_k \ge 0$, $\psi_k \in C_0^\infty(\Omega)$, $\operatorname{supp}(\psi_k) \subset \overline{\Omega_{k+1}} \smallsetminus \Omega_{k-1}$, and

$$\sum_{k=1}^\infty \psi_k(x) = 1 \quad \forall\ x \in \Omega \ .$$

At each $x \in \Omega$, this sum has at most two nonzero terms. (We say that $\{\psi_k\}_{k=1}^\infty$ is a *partition of unity*.)

Now let $\varepsilon > 0$ be given and φ be an approximate identity as in Lemma 6.6. For $f \in W^{m,p}(\Omega)$, choose, by Corollary 6.7, $\varepsilon_k > 0$ small enough that $\varepsilon_k \le \frac{1}{2}\operatorname{dist}(\Omega_{k+1}, \partial\Omega_{k+2})$ and

$$\|\varphi_{\varepsilon_k} * (\psi_k f) - \psi_k f\|_{W^{m,p}} \le \varepsilon 2^{-k} \ .$$

Then $\operatorname{supp}(\varphi_{\varepsilon_k} * (\psi_k f)) \subset \overline{\Omega_{k+2}} \smallsetminus \Omega_{k-2}$, so set

$$g = \sum_{k=1}^\infty \varphi_{\varepsilon_k} * (\psi_k f) \in C^\infty \ ,$$

which is a finite sum at any point $x \in \Omega$, and note that

$$\|f - g\|_{W^{m,p}(\Omega)} \le \sum_{k=1}^\infty \|\psi_k f - \varphi_{\varepsilon_k} * (\psi_k f)\|_{W^{m,p}} \le \varepsilon \sum_{k=1}^\infty 2^{-k} = \varepsilon \ .$$

$\square$

The space $C_0^\infty(\Omega) = \mathcal{D}(\Omega)$ is dense in a generally smaller Sobolev space.

DEFINITION. We let $W_0^{m,p}(\Omega)$ be the closure in $W^{m,p}(\Omega)$ of $C_0^\infty(\Omega)$.

PROPOSITION 6.8. *If* $1 \le p < \infty$, *then*

(a) $W_0^{m,p}(\mathbb{R}^d) = W^{m,p}(\mathbb{R}^d)$,
(b) $W_0^{m,p}(\Omega) \hookrightarrow W^{m,p}(\Omega)$ *(continuously imbedded)*,
(c) $W_0^{0,p}(\Omega) = L_p(\Omega)$.

The dual of $L_p(\Omega)$ is $L_q(\Omega)$, when $1 \le p < \infty$ and $1/p + 1/q = 1$. Since $W^{m,p}(\Omega) \hookrightarrow L_p(\Omega)$, $L_q(\Omega) \subset (W^{m,p}(\Omega))^*$. In general, the dual of $W^{m,p}(\Omega)$ is much larger than $L_q(\Omega)$, and consists of objects that are more general than distributions. We therefore restrict attention here to $W_0^{m,p}(\Omega)$; its dual functionals act on functions with m derivatives, so in essence they "lack" derivatives.

Definition. For $1 \le p < \infty$, $1/p + 1/q = 1$, and $m \ge 0$ an integer, let

$$(W_0^{m,p}(\Omega))^* = W^{-m,q}(\Omega) \ .$$

Proposition 6.9. *If $1 \le p < \infty$ $(1 < q \le \infty)$, $W^{-m,q}(\Omega)$ consists of distributions that have unique, continuous extensions from $\mathcal{D}(\Omega)$ to $W_0^{m,p}(\Omega)$.*

Proof. Note that open sets of $W^{m,p}(\Omega)$ defined by $\|\cdot\|_{m,p,\Omega}$, when restricted to $C_0^\infty(\Omega)$, are also open in $\mathcal{D}(\Omega)$. That is, $\mathcal{D}(\Omega) \hookrightarrow W^{m,p}(\Omega)$, since inclusion $i : \mathcal{D}(\Omega) \to W^{m,p}(\Omega)$ is continuous (the inverse image of an open set in $W^{m,p}(\Omega)$ is open in $\mathcal{D}(\Omega)$). Thus, given $T \in W^{-m,q}(\Omega)$, $T \circ i \in \mathcal{D}'(\Omega)$, so $T \circ i$ has an extension to $W_0^{m,p}(\Omega)$. That this extension is unique is due to Theorem 5.20, since $\mathcal{D}(\Omega)$ is dense in $W_0^{m,p}(\Omega)$. $\square$

Extensions of distributions from $\mathcal{D}(\Omega)$ to $W^{m,p}(\Omega)$ are *not* necessarily unique, since $\mathcal{D}(\Omega)$ is not necessarily dense. Thus $(W^{m,p}(\Omega))^*$ may contain objects that are *not* distributions.

6.2. Extensions from Ω to $\mathbb{R}^d$

If $\Omega \subsetneqq \mathbb{R}^d$, how are $W^{m,p}(\Omega)$ and $W^{m,p}(\mathbb{R}^d)$ related? It would seem plausible that $W^{m,p}(\Omega)$ is exactly the set of restrictions to Ω of functions in $W^{m,p}(\mathbb{R}^d)$. However, the boundary of Ω, $\partial\Omega$, plays a subtle role, and our conjecture is true only for reasonable Ω, as we will see in this section.

The converse to our question is: given $f \in W^{m,p}(\Omega)$, can we find $\tilde f \in W^{m,p}(\mathbb{R}^d)$ such that $f = \tilde f$ on Ω. The existence of such an *extension* $\tilde f$ of f can be very useful.

Lemma 6.10. *If Ω is a half space in $\mathbb{R}^d$, $1 \le p < \infty$, and $m \ge 0$ is fixed, then there is a bounded linear extension operator*

$$E : W^{m,p}(\Omega) \to W^{m,p}(\mathbb{R}^d) \ ;$$

that is, for $f \in W^{m,p}(\Omega)$, $Ef|_\Omega = f$ and there is some $C > 0$ such that

$$\|Ef\|_{W^{m,p}(\mathbb{R}^d)} \le C\|f\|_{W^{m,p}(\Omega)} \ .$$

Note that in fact

$$\|f\|_{W^{m,p}(\Omega)} \le \|Ef\|_{W^{m,p}(\mathbb{R}^d)} \le C\|f\|_{W^{m,p}(\Omega)} \ ,$$

so $\|f\|_{W^{m,p}(\Omega)}$ and $\|Ef\|_{W^{m,p}(\mathbb{R}^d)}$ are comparable.

Proof. Choose a coordinate system so that

$$\Omega = \{x \in \mathbb{R}^d : x_d > 0\} \equiv \mathbb{R}^d_+ \ .$$

If f is defined (almost everywhere) on $\mathbb{R}^d_+$, we extend f to the rest of $\mathbb{R}^d$ by reflection about $x_d = 0$. A simple reflection would not preserve differentiation, so we use the following construction.

For almost every $x \in \mathbb{R}^d$, let

$$Ef(x) = \begin{cases} f(x) & \text{if } x_d > 0, \\ \displaystyle\sum_{j=1}^{m+1} \lambda_j f(x_1, \dots, x_{d-1}, -jx_d) & \text{if } x_d \le 0, \end{cases}$$

where the numbers λ_j are defined below. Clearly E is a linear operator.

If $f \in C^m(\overline{\mathbb{R}^d_+}) \cap W^{m,p}(\mathbb{R}^d_+)$, then for any integer k between 0 and m,

$$D_d^k Ef(x) = \begin{cases} D_d^k f(x_1, \dots, x_{d-1}, x_d) & \text{if } x_d > 0 \\ \displaystyle\sum_{j=1}^{m+1} (-j)^k \lambda_j D_d^k f(x_1, \dots, x_{d-1}, -jx_d) & \text{if } x_d < 0. \end{cases}$$

We claim that we can choose the λ_j such that

$$\sum_{j=1}^{m+1} (-j)^k \lambda_j = 1 \ , \qquad k = 0, 1, \dots, m \ . \tag{6.1}$$

If so, then $D_d^k Ef(x)$ is continuous as $x_d \to 0$, and so $Ef \in C^m(\mathbb{R}^d)$. Thus for $|\alpha| \le m$,

$$\begin{aligned} \|D^\alpha Ef\|^p_{L_p(\mathbb{R}^d)} &= \|D^\alpha f\|^p_{L_p(\mathbb{R}^d_+)} + \int_{\mathbb{R}^d_+} \Big| \sum_{j=1}^{m+1} (-j)^{\alpha_d} \lambda_j D^\alpha f(x_1, \dots, x_{d-1}, jx_d) \Big|^p \, dx \\ &\le C_{m,p} \|D^\alpha f\|^p_{L_p(\mathbb{R}^d_+)} \ . \end{aligned} \tag{6.2}$$

Let now $f \in W^{m,p}(\mathbb{R}^d_+) \cap C^\infty(\mathbb{R}^d_+)$, extended by zero. For $t > 0$, let τ_t be translation by t in the $(-e_d)$-direction:

$$\tau_t f(x) = f(x + te_d) \ .$$

Translation is continuous in $L_p(\mathbb{R}^d)$, so

$$D^\alpha \tau_t f = \tau_t D^\alpha f \xrightarrow{L_p} D^\alpha f \quad \text{as} \quad t \to 0^+ \ .$$

That is,

$$\tau_t f \xrightarrow{W^{m,p}(\mathbb{R}^d_+)} f \ .$$

But $\tau_t f \in C^\infty(\overline{\mathbb{R}^d_+})$, so in fact $C^\infty(\overline{\mathbb{R}^d_+}) \cap W^{m,p}(\mathbb{R}^d_+)$ is dense in $W^{m,p}(\mathbb{R}^d_+)$. Thus (6.2) extends to all of $W^{m,p}(\mathbb{R}^d_+)$.

We must prove that the λ_j satisfying (6.1) can be chosen. Let $x_j = -(j-1)$, and define the $(m+1) \times (m+1)$ matrix M by

$$M_{ij} = x_j^{i-1} \ .$$

Then (6.1) is the linear system

$$M\lambda = e \ ,$$

where λ is the vector of the λ_j's and e is the vector of 1's. Now M is a Vandermonde matrix, and its determinant is known to be

$$\det M = \prod_{1 \le i < j \le m+1} (x_j - x_i) \ .$$

In our case, $\det M \neq 0$, and so the λ_j's exist (uniquely, in fact). $\square$

We can generalize the Lemma through a smooth distortion of the boundary. We first define what we mean by a smooth boundary.

DEFINITION. For integer $m \geq 0$, the bounded domain $\Omega \subset \mathbb{R}^d$ has a $C^{m,1}$-*boundary* (or a *Lipschitz boundary* if $m = 0$) if there exits a finite number of open sets $\Omega_j \subset \mathbb{R}^d$ with the following properties.

(a) $\overline{\Omega_j} \subset\subset \mathbb{R}^d$ and $\partial\Omega \subset \bigcup_j \Omega_j$.
(b) There are functions $\psi_j : \Omega_j \to B_1(0)$ that are one-to-one and onto such that both ψ_j and ψ_j^{-1} are of class $C^{m,1}$, i.e., $\psi_j \in C^{m,1}(\Omega_j)$ and $\psi_j^{-1} \in C^{m,1}(B_1(0))$.
(c) $\psi_j(\Omega_j \cap \Omega) = B^+ \equiv B_1(0) \cap \mathbb{R}^d_+$ and $\psi_j(\Omega_j \cap \partial\Omega) = B^+ \cap \partial\mathbb{R}^d_+$.

That is, $\partial\Omega$ is covered by the Ω_j, Ω_j can be smoothly distorted by ψ_j into a ball with $\partial\Omega$ distorted to the plane $x_d = 0$. Note that $\psi \in C^{m,1}(\Omega)$ means that $\psi \in C^m(\Omega)$ and, for all $|\alpha| = m$, there is some $C > 0$ such that

$$|D^\alpha\psi(x) - D^\alpha\psi(y)| \leq C|x - y| \ \forall\ x, y \in \Omega \ ;$$

that is, $D^\alpha\psi$ is Lipschitz.

THEOREM 6.11. *If $m \geq 0$, $1 \leq p < \infty$, and domain $\Omega \subset \mathbb{R}^d$ has a $C^{m-1,1}$ boundary, then there is a bounded (possibly nonlinear) extension operator*

$$E : W^{m,p}(\Omega) \to W^{m,p}(\mathbb{R}^d) \ .$$

PROOF. If $m = 0$, Ω may be any domain (and we can extend by zero). If $m \geq 1$, let $\{\Omega\}_{j=1}^N$ and $\{\psi_j\}_{j=1}^N$ be as in the definition of a $C^{m-1,1}$ boundary. Let $\Omega_0 \subset\subset \Omega$ be such that

$$\Omega \subset \bigcup_{j=0}^N \Omega_j \ .$$

Let $\{\phi_k\}_{k=1}^M$ be a C^∞ partition of unity subordinate to this covering; that is, $\phi_k \in C^\infty(\mathbb{R}^d)$, $\operatorname{supp}(\phi_k) \subset \Omega_{j_k}$ for some j_k between 0 and N, and

$$\sum_{k=1}^M \phi_k(x) = 1 \qquad \forall\ x \in \Omega \ .$$

Such a partition is relatively easy to construct (see, e.g., [**Ad**] or [**GT**] for a more general construction). Then for $f \in W^{m,p}(\Omega) \cap C^\infty(\Omega)$, let $f_k = \phi_k f \in W_0^{m,p}(\Omega_{j_k})$. Let E_0 be the extension operator given in the lemma. If $j_k \neq 0$,

$$E_0(f_k \circ \psi_{j_k}^{-1}) \in W_0^{m,p}(B_1(0)) \ ,$$

so

$$E_0(f_k \circ \psi_{j_k}^{-1}) \circ \psi_{j_k} \in W_0^{m,p}(\Omega_{j_k}) \ .$$

Extend this by zero to all of $\mathbb{R}^d$. We define E by

$$Ef = \sum_{\substack{k=1 \\ (j_k=0)}}^M \phi_k f + \sum_{\substack{k=1 \\ (j_k\neq0)}}^M E_0\Big((\phi_k f) \circ \psi_{j_k}^{-1}\Big) \circ \psi_{j_k} \in W_0^{m,p}\Big(\bigcup_{j=0}^N \Omega_j\Big) \ .$$

Note that derivatives of Ef are in $L_p(\mathbb{R}^d)$ because the ψ_j and $\psi_j^{-1} \in C^{m-1,1}$ (i.e., derivatives up to order m of ψ_j and ψ_j^{-1} are bounded), and so $Ef \in W^{m,p}(\mathbb{R}^d)$, $Ef|_\Omega = f$, and

$$\|Ef\|_{W^{m,p}(\mathbb{R}^d)} \leq C\|f\|_{W^{m,p}(\Omega)} ,$$

where $C \geq 0$ depends on m, p, and Ω through the Ω_j and ψ_j. $\square$

We remark that if $\bar{\Omega} \subset\subset \tilde{\Omega} \subset \mathbb{R}^d$, then we can assume that $Ef \in W_0^{m,p}(\tilde{\Omega})$. To see this, take any $\phi \in C_0^\infty(\tilde{\Omega})$ with $\phi \equiv 1$ on $\bar{\Omega}$, and define a new bounded extension operator by ϕEf.

Many generalizations of this result are possible. In 1961, Calderón gave a proof assuming only that Ω is Lipschitz. In 1970, Stein [**St**] gave a proof where a single operator E can be used for *any* values of m and p (and Ω is merely Lipschitz). Accepting the extension to Lipschitz domains, we have the following characterization of $W^{m,p}(\Omega)$.

Corollary 6.12. *If Ω has a Lipschitz boundary, $1 \leq p < \infty$, and $m \geq 0$, then*

$$W^{m,p}(\Omega) = \{f|_\Omega : f \in W^{m,p}(\mathbb{R}^d)\} .$$

If we restrict to the $W_0^{m,p}(\Omega)$ spaces, extension by 0 gives a bounded extension operator, even if $\partial\Omega$ is ill-behaved.

Theorem 6.13. *Suppose $\Omega \subset \mathbb{R}^d$, $1 \leq p < \infty$, and $m \geq 0$. Let E be defined on $W_0^{m,p}(\Omega)$ as the operator that extends the domain of the function to $\mathbb{R}^d$ by 0; that is, for $f \in W_0^{m,p}(\Omega)$,*

$$Ef(x) = \begin{cases} f(x) & \text{if } x \in \Omega, \\ 0 & \text{if } x \notin \Omega. \end{cases}$$

Then $E : W_0^{m,p}(\Omega) \to W^{m,p}(\mathbb{R}^d)$.

Of course, then

$$\|f\|_{W_0^{m,p}(\Omega)} = \|Ef\|_{W^{m,p}(\mathbb{R}^d)} .$$

Proof. If $f \in W_0^{m,p}(\Omega)$, then there is a sequence $\{f_j\}_{j=1}^\infty \subset C_0^\infty(\Omega)$ such that

$$f_j \xrightarrow{W^{m,p}(\Omega)} f .$$

Let $\phi \in \mathcal{D}$. Then as distributions for $|\alpha| \leq m$,

$$\begin{aligned} \int_\Omega D^\alpha f \phi \, dx \leftarrow \int_\Omega D^\alpha f_j \phi \, dx &= (-1)^{|\alpha|} \int_\Omega f_j \, D^\alpha \phi \, dx \\ &\to (-1)^{|\alpha|} \int_\Omega f \, D^\alpha \phi \, dx \\ &= (-1)^{|\alpha|} \int_{\mathbb{R}^d} Ef \, D^\alpha \phi \, dx \\ &= \int_{\mathbb{R}^d} D^\alpha Ef \, \phi \, dx , \end{aligned}$$

so $ED^\alpha f = D^\alpha Ef$ in $\mathcal{D}'$. The former is an $L_{1,\text{loc}}$ function on $\mathbb{R}^d$, so the Lebesgue Lemma (Prop. 4.7) implies that the two agree as functions. Thus

$$\|f\|_{W^{m,p}(\Omega)} = \left\{ \sum_{|\alpha| \leq m} \int_{\mathbb{R}^d} |ED^\alpha f|^p \, dx \right\}^{1/p} = \|Ef\|_{W^{m,p}(\mathbb{R}^d)} . \qquad \square$$

6.3. The Sobolev Imbedding Theorem

A measurable function f fails to lie in some L_p space either because it blows up or its tail fails to converge to 0 fast enough (consider $|x|^{-\alpha}$ near 0 or for $|x| > R > 0$). However, if Ω is bounded and $f \in W^{m,p}(\Omega)$, $m \geq 1$, the derivative is well behaved, so the function cannot blow up as fast as an arbitrary function and we expect $f \in L_q(\Omega)$ for some $q > p$.

EXAMPLE. Consider $\Omega = (0, 1/2)$ and

$$f(x) = \frac{1}{\log x}$$

for which

$$f'(x) = \frac{-1}{x(\log x)^2} \ .$$

The change of variable $y = -\log x$ ($x = e^{-y}$) shows $f \in W^{1,1}(\Omega)$. In fact, $f' \in L_p(\Omega)$ only for $p = 1$. But $f \in L_p(\Omega)$ for any $p \geq 1$.

We give in this section a precise statement to this idea of trading derivatives for bounds in higher index L_p spaces. Surprisingly, if we have enough derivatives, the function will not only lie in L_∞, but it will in fact be continuous. We begin with an important estimate.

THEOREM 6.14 (Sobolev Inequality). *If $1 \leq p < d$ and*

$$q = \frac{dp}{d-p} \ ,$$

then there is a constant $C = C(d,p)$ such that

$$\|u\|_{L_q(\mathbb{R}^d)} \leq C \|\nabla u\|_{L_p(\mathbb{R}^d)} \qquad \forall\ u \in C_0^1(\mathbb{R}^d) \ . \tag{6.3}$$

LEMMA 6.15 (Generalized Hölder). *If $\Omega \subset \mathbb{R}^d$, $1 \leq p_i \leq \infty$ for $i = 1, \dots, m$, and*

$$\sum_{i=1}^m \frac{1}{p_i} = 1 \ ,$$

then for $f_i \in L_{p_i}(\Omega)$, $i = 1, \dots, m$,

$$\int_\Omega f_1(x) \cdots f_m(x)\, dx \leq \|f_1\|_{L_{p_1}(\Omega)} \cdots \|f_m\|_{L_{p_m}(\Omega)} \ .$$

PROOF. The case $m = 1$ is clear. We proceed by induction on m, using the usual Hölder inequality. Let p'_m be conjugate to p_m (i.e., $1/p_m + 1/p'_m = 1$), where we reorder if necessary so $p_m \geq p_i \ \forall\ i < m$. Then

$$\int_\Omega f_1 \cdots f_m\, dx \leq \|f_1 \cdots f_{m-1}\|_{L_{p'_m}} \|f_m\|_{L_{p_m}} \ .$$

Now $p_1/p'_m, \dots, p_{m-1}/p'_m$ lie in the range from 1 to ∞, and

$$\frac{p'_m}{p_1} + \cdots + \frac{p'_m}{p_{m-1}} = 1 \ ,$$

so the induction hypothesis can be applied:

$$\begin{aligned}\|f_1\cdots f_{m-1}\|_{L_{p'_m}} &= \left\{\int |f_1|^{p'_m}\cdots|f_{m-1}|^{p'_m}\,dx\right\}^{1/p'_m}\\ &\le \left\{\left(\int |f_1|^{p_1}\,dx\right)^{p'_m/p_1}\cdots\left(\int |f_{m-1}|^{p_{m-1}}\,dx\right)^{p'_m/p_{m-1}}\right\}^{1/p'_m}\\ &= \|f_1\|_{L_{p_1}}\cdots\|f_{m-1}\|_{L_{p_{m-1}}}\;. \qquad \square\end{aligned}$$

Proof of the Sobolev Inequality. Let $D_i = d/dx_i$, $i = 1,\dots,d$. We begin with the case $p = 1 < d$. For $u \in C_0^1(\mathbb{R}^d)$,

$$|u(x)| = \Big|\int_{-\infty}^{x_i} D_i u(x)\,dx_i\Big| \le \int_{-\infty}^{\infty} |D_i u|\,dx_i \quad \forall\ i\ ,$$

and so

$$|u(x)|^{d/d-1} \le \prod_{i=1}^{d}\left(\int_{-\infty}^{\infty}|D_i u|\,dx_i\right)^{1/d-1}\ .$$

Integrate this over $\mathbb{R}^d$ and use generalized Hölder in each variable separately for $d-1$ functions each with Lebesgue exponent $d-1$. For x_1,

$$\begin{aligned}\int_{\mathbb{R}^d}|u(x)|^{d/d-1}\,dx &\le \int_{\mathbb{R}^d}\prod_{i=1}^{d}\left(\int_{-\infty}^{\infty}|D_i u|\,dx_i\right)^{1/d-1}\,dx\\ &= \int_{\mathbb{R}^{d-1}}\int_{\mathbb{R}}\left(\int_{-\infty}^{\infty}|D_1 u|\,dx_1\right)^{1/d-1}\prod_{i=2}^{d}\left(\int_{-\infty}^{\infty}|D_i u|\,dx_i\right)^{1/d-1}\,dx_1\,dx_2\cdots dx_d\\ &= \int_{\mathbb{R}^{d-1}}\left(\int_{-\infty}^{\infty}|D_1 u|\,dx_1\right)^{1/d-1}\int_{\mathbb{R}}\prod_{i=2}^{d}\left(\int_{-\infty}^{\infty}|D_i u|\,dx_i\right)^{1/d-1}\,dx_1\,dx_2\cdots dx_d\\ &\le \int_{\mathbb{R}^{d-1}}\left(\int_{-\infty}^{\infty}|D_1 u|\,dx_1\right)^{1/d-1}\prod_{i=2}^{d}\left(\int_{\mathbb{R}}\int_{-\infty}^{\infty}|D_i u|\,dx_i\,dx_1\right)^{1/d-1}\,dx_2\cdots dx_2\ .\end{aligned}$$

Continuing for the other variables, we obtain

$$\int_{\mathbb{R}^d}|u(x)|^{d/d-1}\,dx \le \left(\prod_{i=1}^{d}\int_{\mathbb{R}^d}|D_i u|\,dx\right)^{1/d-1}\ .$$

Since for nonnegative numbers $a_1,\dots,a_n$,

$$\prod_{i=1}^{n} a_i \le \frac{1}{n}\left(\sum_{i=1}^{n} a_i\right)^{n}\ ,$$

and

$$\left(\sum_{i=1}^{n} a_i\right)^{2} \le n\sum_{i=1}^{n} a_i^2$$

(i.e., in $\mathbb{R}^n$, $|\mathbf{a}|_{\ell_1} \leq \sqrt{n}|\mathbf{a}|_{\ell_2}$), we see that

$$\begin{aligned}\int_{\mathbb{R}^d} |u(x)|^{d/d-1}\,dx &\leq \left\{\frac{1}{d}\left(\sum_{i=1}^d \int_{\mathbb{R}^d} |D_i u|\,dx\right)^d\right\}^{1/d-1}\\ &= \frac{1}{d^{1/d-1}}\left(\int_{\mathbb{R}^d} |\nabla u|_{\ell_1}\,dx\right)^{d/d-1}\\ &\leq \frac{1}{d^{1/d-1}}\left(\sqrt{d}\int_{\mathbb{R}^D} |\nabla u|_{\ell_2}\,dx\right)^{d/d-1} .\end{aligned}$$

Thus for C_d a constant depending on d,

$$\|u\|_{L_{d/d-1}} \leq C_d \|\nabla u\|_{L_1} . \tag{6.4}$$

For $p \neq 1$, we apply (6.4) to $|u|^\gamma$ for appropriate $\gamma > 0$:

$$\begin{aligned}\big\|\,|u|^\gamma\big\|_{L_{d/d-1}} &\leq \gamma C_d \big\|\,|u|^{\gamma-1}|\nabla u|\,\big\|_{L_1}\\ &\leq \gamma C_d \big\|\,|u|^{\gamma-1}\big\|_{L_{p'}} \|\nabla u\|_{L_p} ,\end{aligned}$$

where $1/p + 1/p' = 1$. We choose γ so that

$$\frac{\gamma d}{d-1} = (\gamma - 1)p' ;$$

that is

$$\gamma = \frac{(d-1)p}{d-p} > 0$$

and so

$$\frac{\gamma d}{d-1} = (\gamma-1)p' = \frac{dp}{d-p} = q .$$

Thus

$$\|u\|_{L_q}^\gamma \leq \gamma C_d \|u\|_{L_q}^{\gamma-1} \|\nabla u\|_{L_p} ,$$

and the result follows. □

We get a better result if $p > d$.

LEMMA 6.16. *If $p > d$, then there is a constant $C = C(d,p)$ such that*

$$\|u\|_{L_\infty(\mathbb{R}^d)} \leq C(\operatorname{diam}(\Omega)^d)^{\frac{1}{d}-\frac{1}{p}} \|\nabla u\|_{L_p(\mathbb{R}^d)} \quad \forall\ u \in C_0^1(\mathbb{R}^d) , \tag{6.5}$$

where $\Omega = \operatorname{supp}(u)$ and $\operatorname{diam}(\Omega)$ is the diameter of Ω.

PROOF. Suppose $u \in C_0^1(\mathbb{R}^d)$. For any unit vector $\mathbf{e}$,

$$u(x) = \int_0^\infty \frac{\partial u}{\partial e}(x - \mathbf{e}r)\,dr = \int_0^\infty \nabla u(x - \mathbf{e}r)\cdot \mathbf{e}\,dr ,$$

so integrate over $\mathbf{e} \in S_1(0)$, the unit sphere:

$$\begin{aligned}d\omega_d\, u(x) &= \int_{S_1(0)}\int_0^\infty \nabla u(x - re)\cdot e\,dr\,d\Theta\\ &= \int_{\mathbb{R}^d} \nabla u(x-y)\cdot\frac{y}{|y|}\,\frac{1}{|y|^{d-1}}\,dy ,\end{aligned}$$

where ω_d is the volume of the unit ball.

Now suppose $\operatorname{supp}(u) \subset B_1(x)$. Then for $1/p + 1/p' = 1$,

$$|u(x)| \le \frac{1}{d\omega_d} \|\nabla u\|_{L_p} \big\| \, |y|^{1-d} \big\|_{L_{p'}(B_1(0))}$$

and

$$\begin{aligned} \big\| \, |y|^{1-d} \big\|_{L_{p'}}^{p'} &= \int_{B_1(0)} |y|^{(1-d)p'} \, dy \\ &= d\omega_d \int_0^1 r^{(1-d)p'+d-1} \, dr \\ &= \frac{d\omega_d}{(1-d)p'+d} \, r^{(1-d)p'+d} \Big|_0^1 < \infty \end{aligned}$$

provided $(1-d)p' + d > 0$, i.e., $p > d$. So there is $C_{d,p} > 0$ such that

$$|u(x)| \le C_{d,p} \|\nabla u\|_{L_p} \; .$$

If $\Omega = \operatorname{supp}(u) \not\subset B_1(0)$, for $x \in \Omega$, consider the change of variable

$$y = \frac{x - \bar{x}}{\operatorname{diam}(\Omega)} \in B_1(0) \; ,$$

where $\bar{x}$ is the average of x on Ω. Apply the result to

$$\tilde{u}(y) = u\big(\operatorname{diam}(\Omega) y + \bar{x} \big) \; . \qquad \square$$

We summarize and extend the two previous results in the following lemma.

LEMMA 6.17. *Let $\Omega \subset \mathbb{R}^d$ and $1 \le p < \infty$.*

(a) *If $1 \le p < d$ and $q = dp/(d-p)$, then there is a constant $C > 0$ independent of Ω such that for all $u \in W_0^{1,p}(\Omega)$,*

$$\|u\|_{L_q(\Omega)} \le C \|\nabla u\|_{L_p(\Omega)} \; . \tag{6.6}$$

(b) *If $p = d$ and Ω is bounded, then there is a constant $C_\Omega > 0$ depending on the measure of Ω such that for all $u \in W_0^{1,d}(\Omega)$,*

$$\|u\|_{L_q(\Omega)} \le C_\Omega \|\nabla u\|_{L_d(\Omega)} \quad \forall \; q < \infty \; , \tag{6.7}$$

where C_Ω depends also on q. Moreover, if $p = d = 1$, $q = \infty$ is allowed.

(c) *If $d < p < \infty$ and Ω is bounded, then there is a constant $C > 0$ independent of Ω such that for all $u \in W_0^{1,p}(\Omega)$,*

$$\|u\|_{L_\infty(\Omega)} \le C \big(\operatorname{diam}(\Omega)^d \big)^{\frac{1}{d} - \frac{1}{p}} \|\nabla u\|_{L_p(\Omega)} \; . \tag{6.8}$$

Moreover, $W_0^{1,p}(\Omega) \subset C(\bar{\Omega})$.

PROOF. For (6.6) and (6.8), we extend (6.3) and (6.5) by density. Note that a sequence in $C_0^\infty(\Omega)$, Cauchy in $W_0^{1,p}(\Omega)$, is also Cauchy in $L_q(\Omega)$ if $1 \le p < d$ and in $C^0(\bar{\Omega})$ if $p > d$, since we can apply (6.3) or (6.5) to the difference of elements of the sequence. Moreover, when $p > d$ and Ω bounded, the uniform limit of continuous functions in $C_0^\infty(\Omega) \subset C(\bar{\Omega})$ is continuous on $\bar{\Omega}$, so $W_0^{1,p}(\Omega) \subset C(\bar{\Omega})$.

Consider (6.7). The case $d = 1$ is a consequence of the Fundamental Theorem of Calculus and left to the reader. Since Ω is bounded, the Hölder inequality implies $L_{p_1}(\Omega) \subset L_{p_2}(\Omega)$ whenever

$p_1 \geq p_2$. Thus if $p = d > 1$ and $u \in W_0^{1,d}(\Omega)$, also $u \in W_0^{1,p^-}(\Omega)$ for any $1 \leq p^- < p = d$. We apply (6.6) to obtain that

$$\|u\|_{L_q(\Omega)} \leq C\|\nabla u\|_{L_{p^-}(\Omega)} \leq C|\Omega|^{(d-p^-)/d}\|\nabla u\|_{L_d(\Omega)}$$

for $q \leq dp^-/(d-p^-)$, which can be made as large as we like by taking p^- close to d. $\square$

COROLLARY 6.18 (Poincaré). *If $\Omega \subset \mathbb{R}^d$ is bounded, $m \geq 0$ and $1 \leq p < \infty$, then the norm on $W_0^{m,p}(\Omega)$ is equivalent to*

$$|u|_{W_0^{m,p}(\Omega)} = \left\{ \sum_{|\alpha|=m} \|D^\alpha u\|_{L_p(\Omega)}^p \right\}^{1/p} .$$

PROOF. Repeatedly use the Sobolev Inequality (6.6) (or (6.7) or (6.8) for larger p) and the fact that $L_q(\Omega) \subset L_p(\Omega)$ for $q \geq p$. $\square$

That is, only the highest order derivatives are needed in the $W_0^{m,p}(\Omega)$-norm. This is an important result that we will use later when studying boundary value problems.

DEFINITION. We let

$$C_B^j(\Omega) = \{u \in C^j(\Omega) : D^\alpha u \in L_\infty(\Omega) \ \forall\ |\alpha| \leq j\} .$$

This is a Banach space containing $C^j(\bar{\Omega})$. We come now to our main result.

THEOREM 6.19 (Sobolev Imbedding Theorem). *Let $\Omega \subset \mathbb{R}^d$, $j \geq 0$ and $m \geq 1$ integers, and $1 \leq p < \infty$. The following continuous imbeddings hold.*

(a) *If $mp \leq d$, then*

$$W_0^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega) \quad \forall \ \text{finite } q \leq \frac{dp}{d-mp}$$

with $q \geq p$ if Ω is unbounded.

(b) *If $mp > d$ and Ω bounded, then*

$$W_0^{j+m,p}(\Omega) \hookrightarrow C_B^j(\Omega) .$$

Moreover, if Ω has a bounded extension operator on $W^{j+m,p}$, or if $\Omega = \mathbb{R}^d$, then the following hold.

(c) *If $mp \leq d$, then*

$$W^{j+m,p}(\Omega) \hookrightarrow W^{j,q}(\Omega) \quad \forall \ \text{finite } q \leq \frac{dp}{d-mp}$$

with $q \geq p$ if Ω unbounded.

(d) *If $mp > d$ then*

$$W^{j+m,p}(\Omega) \hookrightarrow C_B^j(\Omega) .$$

PROOF. We begin with some remarks that simplify our task.

Note that the results for $j = 0$ extend immediately to the case for $j > 0$. We claim the results for $m = 1$ also extend by iteration to the case $m > 1$. The critical exponent q_m that separates case (a) from (b), or (c) from (d), satisfies for $m = 1, 2, \ldots,$

$$q_m = \frac{dp}{d-mp}$$

which implies that for $0 \le k < m$,

$$q_{k+1} = \frac{dp}{d-(k+1)p} = \frac{dq_k}{d-q_k} \ .$$

When we apply the $m = 1$ result successively to a series of Lebesgue exponents, we never change case; thus, we obtain the final result for $m > 1$.

We also claim that the results for $\Omega = \mathbb{R}^d$ imply the results for $\Omega \neq \mathbb{R}^d$ through the bounded extension operator E. If $u \in W^{m,p}(\Omega)$, then $Eu \in W^{m,p}(\mathbb{R}^d)$ and we apply the result to Eu. The boundedness of E allows us to restrict back to Ω. For the $W_0^{m,p}(\Omega)$ spaces, we have E defined by extension by 0 for any domain, so the argument can be applied to this case as well.

We have simplified our task to the case of $\Omega = \mathbb{R}^d$, $m = 1$, and $j = 0$.

Consider the case of $p \le d$, and take any $v \in W^{1,p}(\mathbb{R}^d)$ such that $\|v\|_{W^{1,p}(\mathbb{R}^d)} \le 1$. We wish to apply (6.6) or (6.7) to v. To do so, we must restrict to a bounded domain and lie in $W_0^{1,p}$.

Let $R = (-1,1)^d$ be a cube centered at 0, and $\tilde{R} = (-2,2)^d \supset\supset \bar{R}$. Let $\beta \in \mathbb{Z}^d$ be any vector with integer components. Clearly

$$\mathbb{R}^d = \bigcup_\beta (R+\beta) = \bigcup_\beta (\tilde{R}+\beta)$$

is decomposed into bounded domains; however, $v|_{R+\beta}$ does not lie in $W_0^{1,p}(R+\beta)$. Let

$$E : W^{1,p}(R) \to W_0^{1,p}(\tilde{R})$$

be a bounded extension operator with bounding constant C_E. By translation we define the extension operator

$$E_\beta : W^{1,p}(R+\beta) \to W_0^{1,p}(\tilde{R}+\beta) \ ,$$

i.e., by

$$E_\beta(\psi) = E(\tau_{-\beta}\psi) = E(\psi(\cdot - \beta)) \ .$$

Obviously the bounding constant for E_β is also C_E.

Now we can apply (6.6) or (6.7) to

$$E_\beta(v|_{R+\beta})$$

to obtain, for appropriate q,

$$\|E_\beta(v|_{R+\beta})\|_{L_q(\tilde{R}+\beta)} \le C_S \|\nabla E_\beta(v|_{R+\beta})\|_{L_p(\tilde{R}+\beta)} \ ,$$

where C_S is independent of β. Thus

$$\begin{aligned} \|v\|^q_{L_q(R+\beta)} &\le \|E_\beta(v|_{R+\beta})\|^q_{L_q(\tilde{R}+\beta)} \\ &\le C_S^q \|\nabla E_\beta(v|_{R+\beta})\|^q_{L_q(\tilde{R}+\beta)} \\ &\le C_S^q C_E^q \|v\|^q_{W^{1,p}(R+\beta)} \\ &\le C_S^q C_E^q \|v\|^p_{W^{1,p}(R+\beta)} \ , \end{aligned}$$

since $p \le q$ and $\|v\|_{W^{1,p}(\mathbb{R}^d)} \le 1$. Summing over β gives

$$\|v\|_{L_q(\mathbb{R}^d)} \le C$$

for some $C > 0$, since the union of the $R+\beta$ cover $\mathbb{R}^d$ a finite number of times.

If now $u \in W^{1,p}(\mathbb{R}^d)$, $u \neq 0$, let

$$v = \frac{u}{\|u\|_{W^{1,p}(\mathbb{R}^d)}}$$

to obtain

$$\|u\|_{L_q(\mathbb{R}^d)} \le C\|u\|_{W^{1,p}(\mathbb{R}^d)} \;;$$

thus, (a) and (c) follow.

Finally the argument for $p > d$, i.e., (b) and (d), is similar, since again our bounding constant in (6.8) is independent of β. This completes the proof. $\square$

Remark. The extension operator need only work for $W^{1,p}(\Omega)$, since we iterated the one derivative case. Thus Lipschitz domains satisfy the requirements. Most domains of interest (e.g., any polygon or polytope) have Lipschitz boundaries.

6.4. Compactness

We have an important compactness result for Sobolev spaces.

Theorem 6.20 (Rellich-Kondrachov). *If $\Omega \subset \mathbb{R}^d$ is a bounded domain with a Lipschitz boundary, $1 \le p < \infty$, and $j \ge 0$ and $m \ge 1$ are integers, then $W^{j+m,p}(\Omega)$ and $W_0^{j+m,p}(\Omega)$ are compactly imbedded in $W^{j,q}(\Omega)$ $\forall\ 1 \le q < dp/(d-mp)$ if $mp \le d$, and in $C^j(\bar{\Omega})$ if $mp > d$.*

Proof. (Sketch only — see, e.g., [**Ad**, p. 144–8] or [**GT**, p. 167–8].) We show the result for $W_0^{m,p}(\Omega)$, and use extension to bounded $\tilde{\Omega} \supset \Omega$ for $W^{m,p}(\Omega)$. We show for $j = 0$ and $m = 1$, and iterate for the general result.

If $p > d$, let B be any ball. For $u \in W_0^{1,p}(\Omega)$, let

$$u_B = \frac{1}{|B|}\int_B u(x)\,dx$$

be the average of u on B. In a manner similar to the proof of Lemma 6.16, for a.e. $x \in B$,

$$|u(x) - u_B| \le C\int_B |\nabla u(x-y)|\,|y|^{1-d}\,dy\ .$$

This is enough to show equicontinuity of the functions, and then the Ascoli-Arzelá Theorem implies compactness in $C^0(\bar{\Omega})$.

If $p \le d$, we assume initially that $q = 1$. Let $A \subset W_0^{1,p}(\Omega)$ be a bounded set. By density we may assume $A \subset C_0^1(\Omega)$. We may also assume $\|u\|_{W_0^{1,p}(\Omega)} \le 1\ \forall\ u \in A$. For $\varphi \in C_0^\infty(B_1(0))$ an approximation to the identity and $\varepsilon > 0$, let

$$A_\varepsilon = \{u * \varphi_\varepsilon : u \in A\}\ .$$

we estimate $|u * \varphi_\varepsilon|$ and $|\nabla(u * \varphi_\varepsilon)|$ to see that A_ε is bounded and equicontinuous in $C^0(\bar{\Omega})$, so it is precompact in $C^0(\bar{\Omega})$ by the Ascoli-Arzelá Theorem, and so also precompact in $L_1(\Omega)$. Next, we estimate

$$\int_\Omega |u(x) - u * \varphi_\varepsilon(x)|\,dx \le \varepsilon\int_\Omega |Du|\,dx\ ,$$

so $u * \varphi_\varepsilon$ is uniformly close to u in $L_1(\Omega)$. It follows that A is precompact in $L_1(\Omega)$ as well.

For $1 < q \le dp/(d-p)$, we use Hölder and (6.6) or (6.7) to show

$$\|u\|_{L_q(\Omega)} \le C\|u\|_{L_1(\Omega)}^\lambda \|\nabla u\|_{L_p(\Omega)}^{1-\lambda}\ ,$$

where $\lambda + (1-\lambda)(1/p - 1/d) = 1/q$. Thus boundedness in $W_0^{1,p}(\Omega)$ and convergence in $L_1(\Omega)$ implies convergence in $L_q(\Omega)$. $\square$

COROLLARY 6.21. *If $\Omega \subset \mathbb{R}^d$ is bounded, $1 \le p < \infty$, and $\{u_j\}_{j=1}^\infty \subset W^{j+m,p}(\Omega)$ is a bounded sequence, then there exists a subsequence $\{u_{j_k}\}_{k=1}^\infty \subset \{u_j\}_{j=1}^\infty$ which converges in $W^{j,q}(\Omega)$ for $q < dp/(d-mp)$ if $mp \le d$, and in $C^j(\bar\Omega)$ if $mp > d$.*

This result is often used in the following way. Suppose

$$u_j \xrightarrow{W^{m,p}(\Omega)} u \quad \text{as} \quad j \to \infty \text{ weakly} .$$

Then $\{u_j\}$ is bounded, so there is a subsequence for which

$$u_{j_k} \xrightarrow{W^{m-1,p}(\Omega)} u \quad \text{as} \quad k \to \infty \text{ strongly} .$$

6.5. The H^s Sobolev Spaces

In this section we give an alternate definition of $W^{m,2}(\mathbb{R}^d) = H^m(\mathbb{R}^d)$ which has a natural extension to nonintegral values of m. These fractional order spaces will be useful in the next section on traces.

If $f \in \mathcal{S}(\mathbb{R})$, then

$$\widehat{Df} = i\xi \hat{f} .$$

This is an example of a *multiplier operator* $T : \mathcal{S} \to \mathcal{S}$ defined by

$$T(f) = (m(\xi)\hat{f}(\xi))^\vee ,$$

where $m(\xi)$, called the *symbol* of the operator, is in $C^\infty(\mathbb{R})$ and has polynomial growth. For $T = D$, $m(\xi) = i\xi$. While $i\xi$ is smooth, it is *not* invertible, so D is a troublesome operator. However $T = 1 - D^2$ has

$$((1-D)^2)f)^\wedge = (1+\xi^2)\hat{f}(\xi) ,$$

and $(1+\xi^2)$ is well behaved, even though it involves two derivatives of f. What is the square root of this operator? Let $f, g \in \mathcal{S}$ and compute using the L_2-inner product:

$$(Tf, g) = (\widehat{Tf}, \hat{g}) = ((1+\xi^2)\hat{f}, \hat{g}) = ((1+\xi^2)^{1/2}\hat{f}, (1+\xi^2)^{1/2}\hat{g}) .$$

Thus $T = S^2$ where

$$(Sf)^\wedge = (1+\xi^2)^{1/2}\hat{f}(\xi) ,$$

and S is like D ($S = (1-D^2)^{1/2}$).

We are thus led to consider in $\mathbb{R}^d$ the symbol for $(I - \Delta)^{1/2}$, which is

$$b_1(\xi) = (1 + |\xi|^2)^{1/2} \in \mathcal{S}'(\mathbb{R}^d) .$$

Then $b_1(\xi)$ is like D in $\mathbb{R}^d$. For other order derivatives, we generalize for $s \in \mathbb{R}$ to

$$b_s(\xi) = (1 + |\xi|^2)^{s/2} \in \mathcal{S}'(\mathbb{R}^d) .$$

In fact $b_s(\xi) \in C^\infty(\mathbb{R}^d)$ and all derivatives grow at most polynomially. Thus we can multiply tempered distributions by $b_s(\xi)$ by Proposition 5.31.

DEFINITION. For $s \in \mathbb{R}$, let $\Lambda^s : \mathcal{S}' \to \mathcal{S}'$ be given by

$$(\Lambda^s u)^\wedge(\xi) = (1 + |\xi|^2)^{s/2} \hat{u}(\xi)$$

for all $u \in \mathcal{S}'$. We call Λ^s the *Bessel potential* of order s.

REMARK. If $u \in \mathcal{S}$, then

$$\Lambda^s u(x) = (2\pi)^{-d/2} \check{b}_s * u(x) \ .$$

PROPOSITION 6.22. *For any $s \in \mathbb{R}$, $\Lambda^s : \mathcal{S}' \to \mathcal{S}'$ is a continuous, linear, one-to-one, and onto map. Moreover*

$$\Lambda^{s+t} = \Lambda^s \Lambda^t \quad \forall\ s, t \in \mathbb{R}$$

and

$$(\Lambda^s)^{-1} = \Lambda^{-s} \ .$$

DEFINITION. For $s \in \mathbb{R}$, let

$$H^s(\mathbb{R}^d) = \{u \in \mathcal{S}' : \Lambda^s u \in L_2(\mathbb{R}^d)\} \ ,$$

and for $u \in H^s(\mathbb{R}^d)$, let

$$\|u\|_{H^s} = \|\Lambda^s u\|_{L_2(\mathbb{R}^d)} \ .$$

We note that $H^m(\mathbb{R}^d)$ has been defined previously as $W^{m,2}(\mathbb{R}^d)$. Our definitions will coincide, as we will see.

PROPOSITION 6.23. *For all $s \in \mathbb{R}$, $\|\cdot\|_{H^s}$ is a norm, and for $u \in H^s$,*

$$\|u\|_{H^s} = \|\Lambda^s u\|_{L_2} = \left\{ \int_{\mathbb{R}^d} (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 \, d\xi \right\}^{1/2} \ .$$

Moreover, $H^0 = L_2$.

PROOF. Apply the Plancherel Theorem. □

TECHNICAL LEMMA. *For integer $m \geq 0$, there are constants $C_1, C_2 > 0$ such that*

$$C_1 (1 + x^2)^{m/2} \leq \sum_{k=0}^m x^k \leq C_2 (1 + x^2)^{m/2}$$

for all $x \geq 0$.

PROOF. We need constants $c_1, c_2 > 0$ such that

$$c_1 (1 + x^2)^m \leq \left(\sum_{k=0}^m x^k \right)^2 \leq c_2 (1 + x^2)^m \quad \forall\ x \geq 0 \ .$$

Consider

$$f(x) = \frac{\left(\sum\limits_{k=0}^m x^k \right)^2}{(1 + x^2)^m} \in C^0([0, \infty)) \ .$$

Since $f(0) = 1$ and $\lim_{x\to\infty} f(x) = 1$, $f(x)$ has a maximum on $[0, \infty)$, which gives c_2. Similarly $g(x) = 1/f(x)$ has a maximum, giving c_1. □

THEOREM 6.24. *If $m \geq 0$ is an integer, then*

$$H^m(\mathbb{R}^d) = W^{m,2}(\mathbb{R}^d) \ .$$

PROOF. If $u \in W^{m,2}(\mathbb{R}^d)$, then $D^\alpha u \in L_2 \ \forall \ |\alpha| \leq m$. But then

$$|\xi|^k |\hat{u}(\xi)| \in L_2 \quad \forall \ k \leq m \ ,$$

which is equivalent by the lemma to saying that

$$(1 + |\xi|^2)^{m/2} |\hat{u}(\xi)| \in L_2 \ .$$

That is, $u \in H^m(\mathbb{R}^d)$. For $u \in H^m$, we reverse the steps above to conclude that $u \in W^{m,2}$. Moreover, we have shown that the norms are equivalent. □

PROPOSITION 6.25. *A compatible inner product on $H^s(\mathbb{R}^d)$ for any $s \in \mathbb{R}$ is given by*

$$(u, v)_{H^s} = (\Lambda^s u, \Lambda^s v)_{L_2} = \int \Lambda^s u \overline{\Lambda^s v} \, dx$$

for all $u, v \in H^s(\mathbb{R}^d)$. Moreover, $\mathcal{S} \subset H^s$ is dense, and H^s is a Hilbert space.

PROOF. It is easy to verify that $(u, v)_{H^s}$ is an inner product, and easily

$$\|u\|_{H^2}^2 = (u, u)_{H^s} \quad \forall \ u \in H^s \ .$$

Given $\varepsilon > 0$ and $u \in H^s$, there is $f \in \mathcal{S}$ such that

$$\|(1 + |\xi|^2)^{s/2} \hat{u} - f\|_{L_2} < \varepsilon \ ,$$

since $\mathcal{S}$ is dense in L_2. But

$$g = (1 + |\xi|^2)^{-s/2} f \in \mathcal{S} \ ,$$

so

$$\|u - \check{g}\|_{H^s} = \|(1 + |\xi|^2)^{s/2} (\hat{u} - g)\|_{L_2} < \varepsilon \ ,$$

showing that $\mathcal{S}$ is dense in H^s. Finally, if $\{u_j\}_{j=1}^\infty \subset H^s$ is Cauchy, then

$$f_j = (1 + |\xi|^2)^{s/2} \hat{u}_j$$

gives a Cauchy sequence in L_2. Let $f_j \xrightarrow{L_2} f$ and let

$$g = \Big((1 + |\xi|^2)^{-s/2} f \Big)^\vee \in H^s \ .$$

Then

$$\|u_j - g\|_{H^s} = \|f_j - f\|_{L_2} \to 0$$

as $j \to \infty$. Thus H^s is complete. □

These Hilbert spaces form a one-parameter family $\{H^s\}_{s \in \mathbb{R}}$. They are also nested.

PROPOSITION 6.26. *If $s \geq t$, then $H^s \subset H^t$.*

PROOF. If $u \in H^s$, then

$$\begin{aligned} \|u\|_{H^t}^2 &= \int (1 + |\xi|^2)^t |\hat{u}(\xi)|^2 \, d\xi \\ &\leq \int (1 + |\xi|^2)^s |\hat{u}(\xi)|^2 \, dx = \|u\|_{H^s}^2 \ . \end{aligned}$$

□

We note that the negative index spaces are dual to the positive ones.

Proposition 6.27. *If $s \geq 0$, then we may identify $(H^s)^*$ with H^{-s} by the pairing*

$$\langle u, v\rangle = (\Lambda^s u, \Lambda^{-s} v)_{L_2}$$

for all $u \in H^s$ and $v \in H^{-s}$.

Proof. By the Riesz Theorem, $(H^s)^*$ is isomorphic to H^s by the pairing

$$\langle u, w\rangle = (u, w)_{H^s}$$

for all $u \in H^s$ and $w \in H^s \cong (H^s)^*$. But then

$$v = \Big((1 + |\xi|^2)^s \hat{w}\Big)^{\vee} \in H^{-s}$$

gives a one-to-one correspondence between H^{-s} and H^s. Moreover,

$$\|v\|_{H^{-s}} = \|w\|_{H^s} ,$$

so we have H^{-s} isomorphic to $H^s \cong (H^s)^*$. $\square$

Corollary 6.28. *For all integral m, $H^m = W^{m,2}$.*

Proof. For $m \geq 0$, $W^{-m,2} = (W_0^{m,2})^* = (W^{m,2})^*$, since our domain is all of $\mathbb{R}^d$. $\square$

Finally, let us consider restriction to a domain $\Omega \subset \mathbb{R}^d$.

Definition. If $\Omega \subset \mathbb{R}^d$ is a domain and $s \geq 0$, let

$$H^s(\Omega) = \{u|_\Omega : u \in H^s(\mathbb{R}^d)\} .$$

Moreover, let $H_0^s(\Omega)$ be constructed as follows. Map functions in $C_0^\infty(\Omega)$ to $C_0^\infty(\mathbb{R}^d)$ by extending by zero. Take the closure of this space in $H^s(\mathbb{R}^d)$. Finally, restrict back to Ω. We say more concisely but imprecisely that $H_0^s(\Omega)$ is the completion in $H^s(\mathbb{R}^d)$ of $C_0^\infty(\Omega)$.

Let us elaborate on our definition of $H^s(\Omega)$. Let

$$Z = \{u \in H^s(\mathbb{R}^d) : u|_\Omega = 0\} .$$

Then $Z \subset H^s(\mathbb{R}^d)$ is a closed subspace, so we can define the quotient space

$$H^s(\mathbb{R}^d)/Z = \{x + Z : x \in H^s(\Omega)\} ;$$

that is, for $x \in H^s(\mathbb{R}^d)$, let

$$\hat{x} = x + Z$$

be the *coset* of x, and let $H^s(\mathbb{R}^d)/Z$ be the set of cosets (or *equivalence classes* where $x, y \in H^s(\mathbb{R}^d)$ are equivalent if $x - y \in Z$, so $\hat{x} = \hat{y}$). Then $H^s(\mathbb{R}^d)/Z$ is a vector space, and a norm is given by

$$\|\hat{x}\|_{H^s(\mathbb{R}^d)/Z} = \inf_{\substack{\tilde{x}\in H^s(\mathbb{R}^d)\\ \tilde{x}\in\hat{x}}} \|\tilde{x}\|_{H^s(\mathbb{R}^d)} = \inf_{z\in Z} \|x + z\|_{H^s(\mathbb{R}^d)} = \|P_Z^\perp x\|_{H^s(\mathbb{R}^d)} ,$$

where $P_Z^\perp$ is $H^s(\mathbb{R}^d)$-orthogonal projection onto $Z^\perp$. We also have an inner product defined by

$$(\hat{x}, \hat{y})_{H^s(\mathbb{R}^d)/Z} = \tfrac{1}{4}\{\|\hat{x} + \hat{y}\|^2_{H^s(\mathbb{R}^d)/Z} - \|\hat{x} - \hat{y}\|^2_{H^s(\mathbb{R}^d)/Z}\} ,$$

wherein we assume the field $\mathbb{F} = \mathbb{R}$. Moreover, $H^s(\mathbb{R}^d)/Z$ is a Hilbert space. We leave these facts for the reader to verify. Now define

$$\pi : H^s(\mathbb{R}^d)/Z \to H^s(\Omega)$$

by

$$\pi(\hat{x}) = \pi(x + Z) = x|_\Omega \ .$$

This map is well defined, since if $\hat{x} = \hat{y}$, then $x|_\Omega = y|_\Omega$. Moreover, π is linear, one-to-one, and onto. So we define for $x, y \in H^s(\Omega)$

$$\|x\|_{H^s(\Omega)} = \|\pi^{-1}(x)\|_{H^s(\mathbb{R}^d)/Z} = \inf_{\substack{\tilde{x} \in H^s(\mathbb{R}^d) \\ \tilde{x}|_\Omega = x}} \|\tilde{x}\|_{H^s(\mathbb{R}^d)} \ ,$$

and

$$(x, y)_{H^s(\Omega)} = (\pi^{-1}(x), \pi^{-1}(y))_{H^s(\mathbb{R}^d)/Z} = \tfrac{1}{4}\{\|x + y\|^2_{H^s(\Omega)} - \|x - y\|^2_{H^s(\Omega)}\} \ ,$$

so that $H^s(\Omega)$ becomes a Hilbert space.

PROPOSITION 6.29. *If $\Omega \subset \mathbb{R}^d$ is a domain and $s \geq 0$, then $H^s(\Omega)$ is a Hilbert space. Moreover, for any constant $C > 1$, given $u \in H^s(\Omega)$, there is $\tilde{u} \in H^s(\mathbb{R}^d)$ such that*

$$C\|u\|_{H^s(\Omega)} \geq \|\tilde{u}\|_{H^s(\mathbb{R}^d)} \ .$$

If $s = m$ is an integer, then we had previously defined $H^m(\Omega)$ as $W^{m,2}(\Omega)$. If Ω has a Lipschitz boundary, the two definitions coincide, with equivalent, but not equal, norms. This can be seen by considering the bounded extension operator

$$E : W^{m,2}(\Omega) \to W^{m,2}(\mathbb{R}^d) \ ,$$

for which $u \in W^{m,2}(\Omega)$ implies

$$\|Eu\|_{W^{m,2}(\mathbb{R}^d)} \leq C\|u\|_{W^{m,2}(\Omega)} \leq C\|Eu\|_{W^{m,2}(\mathbb{R}^d)} \ .$$

Since $W^{m,2}(\mathbb{R}^d)$ is the same as $H^m(\mathbb{R}^d)$, with equivalent norms,

$$\begin{aligned}
\|u\|_{H^m(\Omega)} &= \inf_{\substack{v \in H^m(\mathbb{R}^d) \\ v|_\Omega = u}} \|v\|_{H^m(\mathbb{R}^d)} \\
&\leq \|Eu\|_{H^m(\mathbb{R}^d)} \leq C_1\|Eu\|_{W^{m,2}(\mathbb{R}^d)} \\
&\leq C_2\|u\|_{W^{m,2}(\Omega)} \leq C_2 \inf_{\substack{v \in W^{m,2}(\mathbb{R}^d) \\ v|_\Omega = u}} \|v\|_{W^{m,2}(\mathbb{R}^d)} \\
&\leq C_3 \inf_{\substack{v \in H^m(\mathbb{R}^d) \\ v|_\Omega = u}} \|v\|_{H^m(\mathbb{R}^d)} = C_3\|u\|_{H^m(\Omega)} \ .
\end{aligned}$$

Thus our two definitions of $H^m(\Omega)$ are consistent, and, depending on the norm used, the constant in the previous proposition may be different than described (i.e., not necessarily any $C > 1$). Summarizing, we have the following result.

PROPOSITION 6.30. *If $\Omega \subset \mathbb{R}^d$ has a Lipschitz boundary and $m \geq 0$ is an integer, then*

$$H^m(\Omega) = W^{m,2}(\Omega)$$

and the $H^m(\Omega)$ and $W^{m,2}(\Omega)$ norms are equivalent.

6.6. A trace theorem

Given a domain $\Omega \subset \mathbb{R}^d$ and a function $f : \Omega \to \mathbb{R}$, the *trace* of f is its value on the boundary of Ω; i.e., the trace is $f|_{\partial\Omega}$, provided this makes sense. We give a precise meaning and construction when f belongs to a Sobolev space.

We begin by restricting functions to lower dimensional hypersurfaces. Let $0 < k < d$ be an integer, and decompose

$$\mathbb{R}^d = \mathbb{R}^{d-k} \times \mathbb{R}^k \ .$$

If $\phi \in C^0(\mathbb{R}^d)$, then the restriction map

$$R : C^0(\mathbb{R}^d) \to C^0(\mathbb{R}^{d-k})$$

is defined by

$$R\phi(x') = \phi(x', 0) \quad \forall\ x' \in \mathbb{R}^{d-k} \ ,$$

wherein $0 \in \mathbb{R}^k$.

Theorem 6.31. *Let k and d be integers with $0 < k < d$. The restriction map R extends to a bounded linear map from $H^s(\mathbb{R}^d)$ onto $H^{s-k/2}(\mathbb{R}^{d-k})$, provided that $s > k/2$.*

Proof. Since $\mathcal{S}$ is dense in our two Sobolev spaces, it is enough to consider $u \in \mathcal{S}(\mathbb{R}^d)$ where R is well defined. Let $v = Ru \in \mathcal{S}(\mathbb{R}^{d-k})$.

The Sobolev norm involves the Fourier transform, so we compute for $y \in \mathbb{R}^{d-k}$

$$v(y) = (2\pi)^{-(d-k)/2} \int_{\mathbb{R}^{d-k}} e^{i\eta\cdot y} \hat{v}(\eta)\, d\eta \ .$$

But, with $\xi = (\eta, \zeta) \in \mathbb{R}^{d-k} \times \mathbb{R}^k$, this is

$$\begin{aligned} v(y) = u(y,0) &= (2\pi)^{-d/2} \int_{\mathbb{R}^d} e^{i\xi\cdot(y,0)} \hat{u}(\xi)\, d\xi \\ &= (2\pi)^{-(d-k)/2} \int_{\mathbb{R}^{d-k}} e^{i\eta\cdot y} \left[(2\pi)^{-k/2} \int_{\mathbb{R}^k} \hat{u}(\eta,\zeta)\, d\zeta \right] d\eta \ . \end{aligned}$$

Thus

$$\hat{v}(\eta) = (2\pi)^{-k/2} \int_{\mathbb{R}^k} \hat{u}(\eta,\zeta)\, d\zeta \ .$$

Introduce $(1 + |\eta|^2 + |\zeta|^2)^{s/2}(1 + |\eta|^2 + |\zeta|^2)^{-s/2}$ into the integral above and apply Hölder's inequality to obtain

$$|\hat{v}(\eta)|^2 \le (2\pi)^{-k} \int_{\mathbb{R}^k} |\hat{u}(\eta,\zeta)|^2 (1 + |\eta|^2 + |\zeta|^2)^s\, d\zeta \int_{\mathbb{R}^k} (1 + |\eta|^2 + |\zeta|^2)^{-s}\, d\zeta \ .$$

The second factor on the right is

$$\int_{\mathbb{R}^k} (1 + |\eta|^2 + |\zeta|^2)^{-s}\, d\zeta = k\omega_k \int_0^\infty (1 + |\eta|^2 + r^2)^{-s} r^{k-1}\, dr \ .$$

With the change of variable

$$(1 + |\eta|^2)^{1/2} \rho = r \ ,$$

this is

$$k\omega_k (1 + |\eta|^2)^{\frac{k}{2}-s} \int_0^\infty (1 + \rho^2)^{-s} \rho^{k-1}\, d\rho \ ,$$

which is finite provided $-2s+k-1<-1$, i.e., $s>k/2$. Combining, we have shown that there is a constant $C>0$ such that

$$|\hat{v}(\eta)|^2(1+|\eta|^2)^{s-\frac{k}{2}} \leq C^2 \int_{\mathbb{R}^k} |\hat{u}(\eta,\zeta)|^2(1+|\eta|^2+|\zeta|^2)^s \, d\zeta \ .$$

Integrating in η gives the bound

$$\|v\|_{H^{s-k/2}(\mathbb{R}^{d-k})} \leq C\|u\|_{H^s(\mathbb{R}^d)} \ .$$

Thus R is a bounded linear operator mapping into $H^{s-k/2}(\mathbb{R}^{d-k})$.

To see that R maps onto $H^{s-k/2}(\mathbb{R}^{d-k})$, let $v \in \mathcal{S}(\mathbb{R}^{d-k})$ and extend v to $\tilde{u} \in C^\infty(\mathbb{R}^d)$ by

$$\tilde{u}(y,z) = v(y) \quad \forall \ y \in \mathbb{R}^{d-k} \ , \ z \in \mathbb{R}^k \ .$$

Now let $\psi \in C_0^\infty(\mathbb{R}^k)$ be such that $\psi(z)=1$ for $|z|<1$ and $\psi(z)=0$ for $|z|>2$. Then

$$u(y,z) = \psi(z)\tilde{u}(y,z) \in \mathcal{S}(\mathbb{R}^d) \ ,$$

and $Ru=v$. ☐

Remark. We saw in the Sobolev Imbedding Theorem that

$$H^s(\mathbb{R}^d) \hookrightarrow C_B^0(\mathbb{R}^d)$$

for $s>d/2$. Thus we can even restrict to a point ($k=d$ above).

If $\Omega \subset \mathbb{R}^d$ has a bounded extension operator $E : H^s(\Omega) \to H^s(\mathbb{R}^d)$, and if $\partial\Omega$ is $C^{m,1}$ smooth for some $m \geq 0$, we can extend our result. If $\{\Omega_j\}_{j=1}^N$ and ψ_j are as given in the definition of $C^{m,1}$ smooth, Ω_0 is open and $\bar{\Omega}_0 \subset \Omega$, and $\Omega \subset \bigcup_{j=0}^N \Omega_j$, let $\{\phi_k\}_{k=1}^M$ be a C^∞ partition of unity subordinate to the cover, so $\operatorname{supp}(\phi_k) \subset \Omega_{j_k}$ for some j_k. Then define for $u \in H^s(\Omega)$,

$$u_k = E(\omega_k u) \circ \psi_{j_k}^{-1} : B_1(0) \to \mathbb{F} \ ,$$

so $u_k \in H_0^s(B_1(0))$ provided $m+1 \geq s$. We restrict u to $\partial\Omega$ by restricting u_k to $S \equiv B_1(0) \cap \{x_d = 0\}$. Since $\operatorname{supp}(u_k) \subset\subset B_1(0)$, we can extend by zero and apply Theorem 6.31 to obtain

$$\|u_k\|_{H^{s-1/2}(S)} \leq C_1 \|u_k\|_{H^s(B_1(0))} \ .$$

We need to combine the u_k and change variables back to Ω and $\partial\Omega$.

It is possible to continue for general $s>0$; however, the technical details become intense. We will instead restrict to integral $m>0$, as this is the case used in the next chapter.

Summing on k, we obtain

$$\sum_{k=1}^M \|u_k\|_{H^{m-1/2}(S)}^s \leq C_1 \sum_{k=1}^M \|u_k\|_{H^m(B_1(0))}^2 \leq C_2 \sum_{k=1}^M \|(\omega_k u) \circ \psi_{j_k}^{-1}\|_{H^m(\mathbb{R}^d_+)}^2 \ ,$$

using the bound on E. Since m is an integer, the final norm merely involves L_2 norms of (weak) derivatives of $(\omega_k u) \circ \psi^{-1}$. The Leibniz rule, Chain rule, and change of variables imply that each such norm is bounded by the $H^m(\Omega)$ norm of u:

$$\sum_{k=1}^M \|u_k\|_{H^{m-1/2}(S)}^2 \leq C_2 \sum_{k=1}^M \|(\omega_k u) \circ \psi_{j_k}^{-1}\|_{H^m(\mathbb{R}^d_+)}^2 \leq C_3 \|u\|_{H^m(\Omega)}^2 \ .$$

Let the *trace* of u, $\gamma_0 u$, be defined for a.e. $x \in \partial\Omega$ by

$$\gamma_0 u(x) = \sum_{k=1}^M \Big(E(\omega_k u) \circ \psi_{j_k}^{-1} \Big) \Big(\psi_{j_k}(x) \Big) \ .$$

Then we clearly have after change of variable

$$\|\gamma_0 u\|^2_{L^2(\partial\Omega)} \le C_4 \sum_{k=1}^{M} \|u_k\|^2_{L^2(S)} \le C_5 \sum_{k=1}^{M} \|u_k\|^2_{H^{m-1/2}(S)} \le C_6 \|u\|^2_{H^m(\Omega)} \ . \tag{6.9}$$

That is, for $u \in H^m(\Omega)$, we can define its trace $\gamma_0 u$ on $\partial\Omega$ as a function in $L_2(\partial\Omega)$; $\gamma_0 : H^m(\Omega) \to L_2(\partial\Omega)$ is a well defined, bounded linear operator. (The above computations carry over to nonintegral m, as can be seen by using the equivalent norms of the next section. Since we do not prove that those norms are indeed equivalent, we have restricted to integral m.)

Let

$$Z = \{u \in H^m(\Omega) : \gamma_0 u = 0 \ \text{ on } \ \partial\Omega\} \ ;$$

this set is well defined by (6.9), and is in fact closed in $H^m(\Omega)$. We therefore define

$$H^{m-1/2}(\partial\Omega) = \{\gamma_0 u \in L_2(\partial\Omega) : u \in H^m(\Omega)\} \ ,$$

which is isomorphic to $H^m(\Omega)/Z$. While $H^{m-1/2}(\partial\Omega) \subset L_2(\partial\Omega)$, we expect that such functions are in fact smoother. A norm is given by

$$\|u\|_{H^{m-1/2}(\partial\Omega)} = \inf_{\substack{\tilde{u} \in H^m(\Omega) \\ \gamma_0 \tilde{u} = u}} \|\tilde{u}\|_{H^m(\Omega)} \ . \tag{6.10}$$

Note that this construction gives immediately the trace theorem

$$\|\gamma_0 u\|_{H^{m-1/2}(\partial\Omega)} \le C \|u\|_{H^m(\Omega)} \ ,$$

where $C = 1$. If an equivalent norm is used for $H^{m-1/2}(\partial\Omega)$, $C \ne 1$ is likely. While we do not have a constructive definition of $H^{m-1/2}(\partial\Omega)$ and its norm that allow us to see explicitly the smoothness of such functions, by analogy to Theorem 6.31 for $\Omega = \mathbb{R}^d_+$, we recognize that $H^{m-1/2}(\partial\Omega)$ functions have intermediate smoothness. The equivalent norm of the next section gives a constructive sense to this statement. We summarize our results.

THEOREM 6.32. *Let $\Omega \subset \mathbb{R}^d$ have a Lipschitz boundary. The trace operator $\gamma_0 : C^0(\bar{\Omega}) \to C^0(\partial\Omega)$ defined by restriction, i.e., $(\gamma_0 u)(x) = u(x)\ \forall\ x \in \partial\Omega$, extends to a bounded linear map*

$$\gamma_0 : H^m(\Omega) \xrightarrow{\text{onto}} H^{m-1/2}(\partial\Omega)$$

for any integer $m \ge 1$.

We can extend this result to higher order derivatives. Tangential derivatives of $\gamma_0 u$ are well defined, since if D_τ is any derivative in a direction tangential to $\partial\Omega$, then

$$D_\tau \gamma_0 u = D_\tau E u = E D_\tau u = \gamma_0 D_\tau u \ .$$

However, derivatives normal to $\partial\Omega$ are more delicate.

DEFINITION. Let $\nu \in \mathbb{R}^d$ be the unit outward normal vector to $\partial\Omega$. Then for $u \in C^1(\bar{\Omega})$,

$$D_\nu u = \frac{\partial u}{\partial \nu} = \nabla u \cdot \nu \ \text{ on } \ \partial\Omega$$

is the normal derivative of u on $\partial\Omega$. If $j \ge 0$ is an integer, let

$$\gamma_j = D^j_\nu = \frac{\partial^j u}{\partial \nu^j} \ .$$

THEOREM 6.33 (Trace Theorem). *Let $\Omega \subset \mathbb{R}^d$ have a $C^{m-1,1} \cap C^{0,1}$ boundary for some integer $m \geq 0$. The map $\gamma : C^m(\bar{\Omega}) \to (C^0(\partial\Omega))^{m+1}$ defined by*

$$\gamma u = (\gamma_0 u, \gamma_1 u, \dots, \gamma_m u)$$

extends to a bounded linear map

$$\gamma : H^{m+1}(\Omega) \xrightarrow{onto} \prod_{j=0}^{m} H^{m-j+1/2}(\partial\Omega) \ .$$

PROOF. Let $u \in H^{m+1}(\Omega) \cap C^\infty(\bar{\Omega})$, which is dense because of the existence of an extension operator. Then iterate the single derivative result for γ_0:

$$\gamma_0 u \in H^{m+1/2}(\partial\Omega) \ , \quad \gamma_1 u = \gamma_0(\nabla u \cdot \nu) \in H^{m-1/2}(\partial\Omega) \ ,$$
$$\gamma_2 u = \gamma_0(\nabla(\nabla u \cdot \nu) \cdot \nu) \in H^{m-3/2}(\partial\Omega) \ , \quad \text{etc.,}$$

wherein we require $\partial\Omega$ to be smooth eventually so that derivatives of ν can be taken, and wherein we have assumed that the vector field ν on $\partial\Omega$ has been extended locally into Ω (that this can be done follows from the Tubular Neighborhood Theorem from topology).

To see that γ maps onto, take

$$v \in \prod_{j=0}^{m} H^{m-j+1/2}(\partial\Omega) \cap C^\infty(\partial\Omega) \ ,$$

and construct $\tilde{v} \in C^\infty(\bar{\Omega}) \cap H^m(\Omega)$ such that

$$\gamma\tilde{v} = v$$

as follows. If $\partial\Omega \subset \mathbb{R}^{d-1}$ we define $\tilde{v}$ as a polynomial

$$\tilde{v}(x', x_d) = v_0(x') + v_1(x')x_d + \cdots + \frac{1}{m!}v_m(x')x_d^m$$

for $x' \in \mathbb{R}^{d-1}$ and $x_d \in \mathbb{R}$, and then multiply by a smooth test function $\psi(x_d)$ that is identically equal to 1 near $x_d = 0$. If $\partial\Omega$ is curved, we decompose $\partial\Omega$ and map it according to the definition of a $C^{m-1,1}$ boundary, and then apply the above construction. $\square$

Recall that

$$H_0^m(\Omega) = W_0^{m,2}(\Omega)$$

is the closure of $C_0^\infty(\Omega)$ in $W^{m,2}(\Omega)$. Since $\gamma u = 0$ for $u \in C_0^\infty(\Omega)$, the same is true for any $u \in H_0^m(\Omega)$. That is, u and its $m-1$ derivatives (normal and/or tangential) vanish on $\partial\Omega$.

THEOREM 6.34. *If $m \geq 1$ is an integer and $\Omega \subset \mathbb{R}^d$ has a $C^{m-1,1}$ boundary, then*

$$H_0^m(\Omega) = \{u \in H^m(\Omega) : \gamma u = 0\} = \ker(\gamma) \ .$$

PROOF. As mentioned above, $H_0^m(\Omega) \subset \ker(\gamma)$. We need to show the opposite inclusion. Again, by a mapping argument of the $C^{m-1,1}$ boundary, we need only consider the case $\Omega = \mathbb{R}_+^d$. Let

$$u \in \ker(\gamma) \cap C_0^\infty(\mathbb{R}^d) \ ,$$

we saw earlier that $C_0^\infty(\mathbb{R}^d)$ is dense in $H^m(\mathbb{R}_+^d)$. Let $\psi \in C^\infty(\mathbb{R})$ be such that $\psi(t) = 1$ for $t > 2$ and $\psi(t) = 0$ for $t < 1$. For $j \geq 1$, let

$$\psi_n(t) = \psi(nt) \ ,$$

which converges to 1 on $\{t > 0\}$ as $n \to \infty$. Then $\psi_n(x_d)u(x) \in C_0^\infty(\mathbb{R}^d_+)$. We claim that

$$\psi_n(x_d)u(x) \xrightarrow{H^m(\mathbb{R}^d_+)} u(x) \quad \text{as} \quad n \to \infty \ .$$

If so, then $u \in H_0^m(\mathbb{R}^d_+)$ as desired.

Let $\alpha \in \mathbb{Z}^d$ be a multi-index such that $|\alpha| \le m$ and let $\alpha = (\beta, \ell)$ where $\beta \in \mathbb{Z}^{d-1}$ and $\ell \ge 0$. Then

$$D^\alpha(\psi_n u - u) = D^\beta D_d^\ell(\psi_n u - u) = \sum_{k=0}^{\ell} \binom{\ell}{k} D_d^{\ell-k}(\psi_n - 1) D^\beta D_d^k u \ ,$$

and we need to show that this tends to 0 in $L_2(\mathbb{R}^d_+)$ as $n \to \infty$. It is enough to show this for each

$$D_d^{\ell-k}(\psi_n - 1) D^\beta D_d^k u = n^{\ell-k} D_d^{\ell-k}(\psi - 1)|_{nx_d} D^\beta D_d^k u \ ,$$

which is clear if $k = \ell$, since the measure of $\{x : \psi_n(x) - 1 > 0\}$ tends to 0. If $k < \ell$, our expression is supported in $\{x \in \mathbb{R}^d_+ : \frac{1}{n} < x_d < \frac{2}{n}\}$, so

$$\|D_d^{\ell-k}(\psi_n - 1) D^\beta D_d^k u\|^2_{L_2(\mathbb{R}^d_+)} \le C_1 n^{2(\ell-k)} \int_{\mathbb{R}^{d-1}} \int_{1/n}^{2/n} |D^\beta D_d^k u(x', x_d)|^2 \, dx_d \, dx' \ .$$

Taylor's theorem implies that for $x = (x', x_d) \in \mathbb{R}^d$ and $j \le m$,

$$\begin{aligned} D_d^k u(x', x_d) = D_d^k u(x', 0) + \cdots + \frac{1}{(j-k-1)!} D_d^{j-1} u(x', 0) \\ + \frac{1}{(j-k-1)!} \int_0^{x_d} (x_d - t)^{j-k-1} D_d^j u(x', t) \, dt \ , \end{aligned}$$

which reduces to the last term since $\gamma u = 0$. Thus for $j = m - |\beta| = \ell \le m$,

$$\begin{aligned} &\|D_d^{\ell-k}(\psi_n - 1) D^\beta D_d^k u\|^2_{L_2(\mathbb{R}^d_+)} \\ &\quad \le C_2 n^{2(\ell-k)} \int_{\mathbb{R}^{d-1}} \int_{1/n}^{2/n} \left| \int_0^{x_d} (x_d - t)^{\ell-k-1} D^\beta D_d^\ell u(x', t) \, dt \right|^2 dx_d \, dx' \\ &\quad \le C_3 n^{2(\ell-k)} n^{-2(\ell-k-1)} \int_{\mathbb{R}^{d-1}} \int_{1/n}^{2/n} \left(\int_0^{2/n} |D^\beta D_d^\ell u(x', t)| \, dt \right)^2 dx_d \, dx' \\ &\quad \le C_4 n^2 \int_{\mathbb{R}^{d-1}} \frac{1}{n^2} \int_0^{2/n} |D^\beta D_d^\ell u(x', t)|^2 \, dt \, dx' \\ &\quad \to 0 \quad \text{as} \quad n \to \infty \end{aligned}$$

since the measure of the inner integral tends to 0. Thus the claim is established and the proof is complete. ◻

6.7. The $W^{s,p}(\Omega)$ Sobolev Spaces

We can generalize some of the $L_2(\Omega)$ results of the last two sections to $L_p(\Omega)$, and the results for integral numbers of derivatives to nonintegral. We summarize a few of the important results. See [**Ad**] for details and precise statements.

DEFINITION. Suppose $\Omega \subset \mathbb{R}^d$, $1 \le p \le \infty$, and $s > 0$ such that $s = m + \sigma$ where $0 < \sigma < 1$ and m is an integer. Then we define for a smooth function u,

$$\|u\|_{W^{s,p}(\Omega)} = \left\{ \|u\|^p_{W^{m,p}(\Omega)} + \sum_{|\alpha|=m} \int_\Omega \int_\Omega \frac{|D^\alpha u(x) - D^\alpha u(y)|^p}{|x-y|^{d+\sigma p}}\, dx\, dy \right\}^{1/p}$$

if $p < \infty$, and otherwise

$$\|u\|_{W^{s,\infty}(\Omega)} = \max\left\{ \|u\|_{W^{m,\infty}(\Omega)}, \max_{|\alpha|=m} \operatorname*{ess\,sup}_{x,y\in\Omega} \frac{|D^\alpha u(x) - D^\alpha u(y)|}{|x-y|^\sigma} \right\} .$$

PROPOSITION 6.35. *For any* $1 \le p \le \infty$, $\|\cdot\|_{W^{s,p}(\Omega)}$ *is a norm.*

DEFINITION. We let $W^{s,p}(\Omega)$ be the completion of $C^\infty(\Omega)$ under the $\|\cdot\|_{W^{s,p}(\Omega)}$-norm, and $W^{s,p}_0(\Omega)$ is the completion of $C^\infty_0(\Omega)$.

PROPOSITION 6.36. *If* $\Omega = \mathbb{R}^d$ *or* Ω *has a Lipschitz boundary, then*

$$W^{s,2}(\Omega) = H^s(\Omega) \quad \text{and} \quad W^{s,2}_0(\Omega) = H^s_0(\Omega) .$$

Thus we have an equivalent norm on $H^s(\Omega)$ *given above.*

If $1 < p < \infty$ and $m = s$ is nonintegral, then we have analogues of the Sobolev Imbedding Theorem, the Rellich-Kondrachov Theorem, and the Trace Theorem. For the Trace Theorem, every time a trace is taken on a hypersurface of one less dimension (as from Ω to $\partial\Omega$), $1/p$ derivative is lost, rather than $1/2$.

6.8. Exercises

1. Prove that for $f \in H^1(\mathbb{R}^d)$, $\|f\|_{H^1(\mathbb{R}^d)}$ is equivalent to

$$\left\{ \int_{\mathbb{R}^d} (1 + |\xi|^2)|\hat{f}(\xi)|^2\, d\xi \right\}^{1/2} .$$

   Can you generalize this to $H^k(\mathbb{R}^d)$?

2. Prove that if $f \in H^1_0(0,1)$, then there is some constant $C > 0$ such that

$$\|f\|_{L_2(0,1)} \le C\|f'\|_{L_2(0,1)} .$$

   If instead $f \in \{g \in H^1(0,1) : \int_0^1 g(x)\, dx = 0\}$, prove a similar estimate.

3. Prove that $\delta_0 \notin (H^1(\mathbb{R}^d))^*$ for $d \ge 2$, but that $\delta_0 \in (H^1(\mathbb{R}))^*$. You will need to define what δ_0 applied to $f \in H^1(\mathbb{R})$ means.

4. Prove that $H^1(0,1)$ is continuously imbedded in $C_B(0,1)$. Recall that $C_B(0,1)$ is the set of bounded and continuous functions on $(0,1)$.

5. Interpolation inequalities.

   (a) Show that for $f \in H^1(\mathbb{R}^d)$ and $0 \le s \le 1$, $\|f\|_{H^s(\mathbb{R}^d)} \le \|f\|^s_{H^1(\mathbb{R}^d)} \|f\|^{1-s}_{L^2(\mathbb{R}^d)}$. Can you generalize this result to $f \in H^r(\mathbb{R}^d)$ for $r > 0$?

   (b) If $\partial\Omega$ is smooth, show that there is a constant C such that for all $f \in H^1(\Omega)$, $\|f\|_{L^2(\partial\Omega)} \le C\|f\|^{1/2}_{H^1(\Omega)} \|f\|^{1/2}_{L^2(\Omega)}$. [Hint: Show for $d = 1$ on $(0,1)$ by considering

$$f(0)^2 = f(x)^2 - \int_0^x \frac{d}{dx} f(t)^2\, dt .$$

For $d > 1$, flatten out $\partial\Omega$ and use a $(d = 1)$-type proof in the normal direction.]

6. Prove that $H^s(\mathbb{R}^d)$ is imbedded in $C^0_B(\mathbb{R}^d)$ if $s > d/2$ by completing the following outline.

   (a) Show that $\displaystyle\int_{\mathbb{R}^d} (1 + |\xi|^2)^{-s}\, d\xi < \infty$.

   (b) If $\phi \in \mathcal{S}$ and $x \in \mathbb{R}^d$, write $\phi(x)$ as the Fourier inversion integral of $\hat{\phi}$. Introduce $1 = (1+|\xi|^2)^s(1+|\xi|^2)^{-s}$ into the integral and apply Hölder to obtain the result for Schwartz class functions.

   (c) Use density to extend the above result to $H^s(\mathbb{R}^d)$.

7. Suppose that $g = \omega * f$, where $f \in L_2(\mathbb{R})$ and $\hat{\omega}(\xi) = \sqrt{|\xi|}$. Determine s such that $g \in H^s(\mathbb{R})$.

8. Elliptic regularity theory shows that if the domain $\Omega \subset \mathbb{R}^d$ has a smooth boundary and $f \in H^s(\Omega)$, then $-\Delta u = f$ in Ω, $u = 0$ on $\partial\Omega$, has a unique solution $u \in H^{s+2}$. For what values of s will u be continuous? Can you be sure that a fundamental solution is continuous? The answers depend on d.

9. Suppose that $\Omega \subset \mathbb{R}^d$ is a bounded set and $\{U_j\}_{j=1}^N$ is a finite collection of open sets in $\mathbb{R}^d$ that cover the closure of Ω (i.e., $\bar{\Omega} \subset \displaystyle\bigcup_{j=1}^N U_j$). Prove that there exists a finite C^∞ partition of unity in Ω subordinate to the cover. That is, construct $\{\phi_k\}_{k=1}^M$ such that $\phi_k \in C_0^\infty(\mathbb{R}^d)$, $\phi_k \subset U_{j_k}$ for some j_k, and
$$\sum_{k=1}^M \phi_k(x) = 1 \ .$$

10. Suppose that $\Omega \subset \mathbb{R}^d$ is a domain and $\{U_\alpha\}_{\alpha\in\mathcal{I}}$ is a collection of open sets in $\mathbb{R}^d$ that cover Ω (i.e., $\Omega \subset \displaystyle\bigcup_{\alpha\in\mathcal{I}} U_\alpha$), Prove that there exists a *locally finite partition of unity in Ω subordinate to the cover*. That is, there exists a sequence $\{\psi_j\}_{j=1}^\infty \subset C_0^\infty(\mathbb{R}^d)$ such that
    (i) For every K compactly contained in Ω, all but finitely many of the ψ_j vanish on K.
    (ii) Each $\psi_j \geq 0$ and $\displaystyle\sum_{j=1}^\infty \psi_j(x) = 1$ for every $x \in \Omega$.
    (iii) For each j, the support of ψ_j is contained in some U_{α_j}, $\alpha_j \in \mathcal{I}$.
    Hints: Let S be a countable dense subset of Ω (e.g., points with rational coordinates). Consider the countable collection of balls $\mathcal{B} = \{B_r(x) \subset \mathbb{R}^d : r$ is rational, $x \in S$, and $B_r(x) \subset U_\alpha$ for some $\alpha \in \mathcal{I}\}$. Order the balls and construct on $B_j = B_{r_j}(x_j)$ a function $\phi_j \in C_0^\infty(B_j)$ such that $0 \leq \phi_j \leq 1$ and $\phi_j = 1$ on $B_{r_j/2}(x_j)$. Then $\psi_1 = \phi_1$ and $\psi_j = (1 - \phi_1)...(1 - \phi_{j-1})\phi_j$ should work.

11. Suppose that $f_j \in H^2(\Omega)$ for $j = 1, 2, ...$, $f_j \overset{w}{\rightharpoonup} f$ weakly in $H^1(\Omega)$, and $D^\alpha f_j \overset{w}{\rightharpoonup} g_\alpha$ weakly in $L_2(\Omega)$ for all multi-indices α such that $|\alpha| = 2$. Show that $f \in H^2(\Omega)$, $D^\alpha f = g_\alpha$, and $f_j \to f$ strongly in in $H^1(\Omega)$.

12. Suppose that $\Omega \subset \mathbb{R}^d$ and $f_j \overset{w}{\rightharpoonup} f$ and $g_j \overset{w}{\rightharpoonup} g$ weakly in $H^1(\Omega)$. Show that $\nabla(f_j g_j) \to \nabla(fg)$ as a distribution. Find all p in $[1, \infty]$ such that the convergence can be taken weakly in $L_p(\Omega)$.

13. Counterexamples.

(a) No imbedding of $W^{1,p}(\Omega) \hookrightarrow L_q(\Omega)$ for $1 \leq p < d$ and $q > dp/(d-p)$. Let $\Omega \subset \mathbb{R}^d$ be bounded and contain 0, and let $f(x) = |x|^\alpha$. Find α so that $f \in W^{1,p}(\Omega)$ but $f \notin L_q(\Omega)$.

(b) No imbedding of $W^{1,p}(\Omega) \hookrightarrow C_B^0(\Omega)$ for $1 \leq p < d$. Note that in the previous case, f is not bounded. What can you say about which (negative) Sobolev spaces the Dirac mass lies in?

(c) No imbedding of $W^{1,p}(\Omega) \hookrightarrow L_\infty(\Omega)$ for $1 < p = d$. Let $\Omega \subset \mathbb{R}^d = B_R(0)$ and let $f(x) = \log(\log(4R/|x|))$. Show $f \in W^{1,p}(B_R(0))$.

(d) $C^\infty \cap W^{1,\infty}$ is not dense in $W^{1,\infty}$. Show that if $\Omega = (-1,1)$ and $u(x) = |x|$, then $u \in W^{1,\infty}$ but $u(x)$ is not the limit of C^∞ functions in the $W^{1,\infty}$-norm.

CHAPTER 7

Boundary Value Problems

We consider in this chapter certain partial differential equations (PDE's) important in science and engineering. Our equations are posed on a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, where typically d is 1, 2, or 3. We also impose auxiliary conditions on the boundary $\partial\Omega$ of the domain, called *boundary conditions* (BC's). A PDE together with its BC's constitute a *boundary value problem* (BVP). We tacitly assume throughout most of this chapter that the underlying field $\mathbb{F} = \mathbb{R}$.

It will be helpful to make the following remark before we begin. The Divergence Theorem implies that for vector $\psi \in (C^1(\bar\Omega))^d$ and scalar $\phi \in C^1(\bar\Omega)$,

$$\int_\Omega \nabla \cdot (\phi\psi)\, dx = \int_{\partial\Omega} \phi\psi \cdot \nu \, d\sigma(x) \ , \tag{7.1}$$

where ν is the unit outward normal vector (which is defined almost everywhere on the boundary of a Lipschitz domain) and $d\sigma$ is the $(d-1)$-dimensional measure on $\partial\Omega$. Since

$$\nabla \cdot (\phi\psi) = \nabla\phi \cdot \psi + \phi\, \nabla \cdot \psi \ ,$$

we have the integration-by-parts formula in $\mathbb{R}^d$

$$\int_\Omega \phi\, \nabla \cdot \psi \, dx = -\int_\Omega \nabla\phi \cdot \psi \, dx + \int_{\partial\Omega} \phi\psi \cdot \nu \, d\sigma(x) \ . \tag{7.2}$$

By density, we extend this formula immediately to the case where merely $\phi \in H^1(\Omega)$ and $\psi \in (H^1(\Omega))^d$. Note that the Trace Theorem 6.33 gives meaning to the boundary integral.

7.1. Second Order Elliptic Partial Differential Equations

Let $\Omega \subset \mathbb{R}^d$ be some bounded Lipschitz domain. The general second order elliptic PDE in divergence form for the unknown function u is

$$-\nabla \cdot (a\nabla u + bu) + cu = f \qquad \text{in } \Omega \ , \tag{7.3}$$

where a is a $d \times d$ matrix, b is a d-vector, and c and f are functions. To be physically relevant and mathematically well posed, it is often the case that $c \geq 0$, $|b|$ is not too large (in a sense to be made clear later), and the matrix a is uniformly positive definite, as defined below.

Definition. If $\Omega \subset \mathbb{R}^d$ is a domain and $a : \bar\Omega \to \mathbb{R}^{d\times d}$ is a matrix, then a is *positive definite* if for a.e. $x \in \bar\Omega$,

$$\xi^T a(x)\xi > 0 \qquad \forall \xi \in \mathbb{R}^d \ , \ \xi \neq 0 \ ,$$

and a is merely *positive semidefinite* if only $\xi^T a(x)\xi \geq 0$. Moreover, a is *uniformly positive definite* if there is some constant $a_* > 0$ such that for a.e. $x \in \bar\Omega$,

$$\xi^T a(x)\xi \geq a_*|\xi|^2 \qquad \forall \xi \in \mathbb{R}^d \ .$$

We remark that positive definiteness of a insures that

$$a\nabla u \cdot \nabla u \geq 0 \ .$$

The positivity of this term can be exploited mathematically. It is also related to physical principles. In many applications, ∇u is the direction of a force and $a\nabla u$ is the direction of a response. Positive definiteness says that the response is generally in the direction of the force, possibly deflected a bit, but never more than 90°.

7.1.1. Practical examples. We provide some examples of systems governed by (7.3).

EXAMPLE (Steady-state conduction of heat). Let $\Omega \subset \mathbb{R}^3$ be a solid body, $u(x)$ the temperature of the body at $x \in \Omega$, and $f(x)$ an external source or sink of heat energy. The heat flux is a vector in the direction of heat flow, with magnitude given as the amount of heat energy that passes through an infinitesimal planar region orthogonal to the direction of flow divided by the area of the infinitesimal region, per unit time. Fourier's Law of Heat Conduction says that the heat flux is $-a\nabla u$, where $a(x)$, the thermal conductivity of the body, is positive definite. Thus, heat flows generally from hot to cold. Finally, $s(x)$ is the specific heat of the body; it measures the amount of heat energy that can be stored per unit volume of the body per degree of temperature. The physical principle governing the system is energy conservation. If $V \subset \Omega$, then the total heat inside V is $\int_V su\,dx$. Changes in time in this total must agree with the external heat added due to f minus the heat lost due to movement through ∂V; thus,

$$\frac{d}{dt}\int_V su\,dx = \int_V f\,dx - \int_{\partial V} (-a\nabla u)\cdot\nu\,d\sigma(x) \ ,$$

where, as always, ν is the outer unit normal vector. Applying the Divergence Theorem, the last term is

$$\int_{\partial V} a\nabla u\cdot\nu\,d\sigma(x) = \int_V \nabla\cdot(a\nabla u)\,dx \ ,$$

and so, assuming the derivative may be moved inside the integral,

$$\int_V \left(\frac{\partial su}{\partial t} - \nabla\cdot(a\nabla u)\right) dx = \int_V f\,dx \ .$$

This holds for every $V \subset \Omega$ with a reasonable boundary. By a modification of Lebesgue's Lemma, we conclude that, except on a set of measure zero,

$$\frac{\partial(su)}{\partial t} - \nabla\cdot(a\nabla u) = f \ . \tag{7.4}$$

In steady-state, the time derivative vanishes, and we have (7.3) with $b = 0$ and $c = 0$. But suppose that $f(x) = f(u(x), x)$ depends on the temperature itself; that is, the external world will add or subtract heat at x depending on the temperature found there. For example, a room Ω may have a thermostatically controlled heater/air conditioner $f = F(u, x)$. Suppose further that $F(u, x) = c(x)(u_{\text{ref}}(x) - u)$ for some $c \geq 0$ and reference temperature $u_{\text{ref}}(x)$. Then

$$\frac{\partial(su)}{\partial t} - \nabla\cdot(a\nabla u) = c(u_{\text{ref}} - u) \ , \tag{7.5}$$

and, in steady-state, we have (7.3) with $b = 0$ and $f = cu_{\text{ref}}$. Note that if $c \geq 0$ and $u \leq u_{\text{ref}}$, then $F \geq 0$ and heat energy is added, tending to increase u. Conversely, if $u \geq u_{\text{ref}}$, u tends to decrease. In fact, in time, $u \to u_{\text{ref}}$. However, if $c < 0$, we have a potentially unphysical situation, in which hot areas (i.e., $u > u_{\text{ref}}$) tend to get even hotter and cold areas even colder. The steady-state configuration would be to have $u = +\infty$ in the hot regions and $u = -\infty$ in the cold regions! Thus $c \geq 0$ should be demanded on physical grounds (later it will be required on mathematical grounds as well).

Example (The electrostatic potential). Let u be the electrostatic potential, for which the electric flux is $-a\nabla u$ for some a measuring the electrostatic permitivity of the medium Ω. Conservation of charge over an arbitrary volume in Ω, the Divergence Theorem, and the Lebesgue Lemma give (7.3) with $c = 0$ and $b = 0$, where f represents the electrostatic charges.

Example (Steady-state fluid flow in a porous medium). The equations of steady-state flow of a nearly incompressible, single phase fluid in a porous medium are similar to those for the flow of heat. In this case, u is the fluid pressure. Darcy's Law gives the volumetric fluid flux (also called the *Darcy velocity*) as $-a(\nabla u - g\rho)$, where a is the permeability of the medium Ω divided by the fluid viscosity, g is the gravitational vector, and ρ is the fluid density. The total mass in volume $V \subset \Omega$ is $\int_V \rho\,dx$, and this quantity changes in time due to external sources (or sinks, if negative, such as wells) represented by f and *mass* flow through ∂V. The mass flux is given by multiplying the volumetric flux by ρ. That is, with t being time,

$$\begin{aligned}\frac{d}{dt}\int_V \rho\,dx &= \int_V f\,dx - \int_{\partial V} -\rho a(\nabla u - g\rho)\cdot\nu\,d\sigma(x)\\ &= \int_V f\,dx + \int_V \nabla\cdot[\rho a(\nabla u - g\rho)]\,dx\ ,\end{aligned}$$

and we conclude that, provided we can take the time derivative inside the integral,

$$\frac{\partial\rho}{\partial t} - \nabla\cdot[\rho a(\nabla u - g\rho)] = f\ .$$

Generally speaking, $\rho = \rho(u)$ depends on the pressure u through an equation-of-state, so this is a time dependent, nonlinear equation. If we assume steady-state flow, we can drop the first term. We might also simplify the equation-of-state if $\rho(u) \approx \rho_0$ is nearly constant (at least over the pressures being encountered). One choice uses

$$\rho(u) \approx \rho_0 + \gamma(u - u_0)\ ,$$

where γ and u_0 are fixed (note that these are the first two terms in a Taylor approximation of ρ about u_0). Substituting this in the equation above results in

$$-\nabla\cdot\{a[(\rho_0 + \gamma(u - u_0))\nabla u - g(\rho_0 + \gamma(u - u_0))^2]\} = f\ .$$

This is still nonlinear, so a further simplification would be to linearize the equation (i.e., assume $u \approx u_0$ and drop all higher order terms involving $u - u_0$). Since $\nabla u = \nabla(u - u_0)$, we obtain finally

$$-\nabla\cdot\{\rho_0 a[\nabla u - g(\rho_0 + 2\gamma(u - u_0))]\} = f\ ,$$

which is (7.3) with a replaced by $\rho_0 a$, $c = 0$, $b = -2\rho_0 a g\gamma$, and f replaced by $f - \nabla\cdot[\rho_0 a g(\rho_0 - 2\gamma u_0)]$.

7.1.2. Boundary conditions (BC's). In each of the previous examples, we determined the equation governing the behavior of the system, given the external forcing term f distributed over the domain Ω. However, the description of each system is incomplete, since we must also describe the external interaction with the world through its boundary $\partial\Omega$.

These boundary conditions generally take one of three forms, though many others are possible depending on the system being modeled. Let $\partial\Omega$ be decomposed into Γ_D, Γ_N, and Γ_R, where the three parts of the boundary are open, contained in $\partial\Omega$, cover $\partial\Omega$ (i.e., $\partial\Omega = \bar\Gamma_D \cup \bar\Gamma_N \cup \bar\Gamma_R$), and are mutually disjoint (so $\Gamma_D \cap \Gamma_N = \Gamma_D \cap \Gamma_R = \Gamma_N \cap \Gamma_R = \emptyset$). We specify the boundary conditions as

$$u = u_D \qquad \text{on } \Gamma_D\ , \tag{7.6}$$

$$-(a\nabla u + bu)\cdot\nu = g_N \qquad \text{on } \Gamma_N\ , \tag{7.7}$$

$$-(a\nabla u + bu)\cdot\nu = g_R(u - u_R) \qquad \text{on } \Gamma_R\ , \tag{7.8}$$

where u_D, u_R, g_N, and g_R are functions with $g_R > 0$. We call (7.6) a *Dirichlet* BC, (7.7) a *Neumann* BC, and (7.8) a *Robin* BC.

The Dirichlet BC fixes the value of the (trace of) the unknown function. In the heat conduction example, this would correspond to specifying the temperature on Γ_D.

The Neumann BC fixes the normal component of the flux $-(a\nabla u + bu)\cdot\nu$. The PDE controls the tangential component, as this component of the flux does not leave the domain in an infinitesimal sense. However, the normal component is the flux into or out of the domain, and so it may be fixed in certain cases. In the heat conduction example, $g_N = 0$ would represent a perfectly insulated boundary, as no heat flux may cross the boundary. If instead heat is added to (or taken away from) the domain through some external heater (or refrigerator), we would specify this through nonzero g_N.

The Robin BC is a combination of the first two types. It specifies that the flux is proportional to the deviation of u from u_R. If $u = u_R$, there is no flux; otherwise, the flux tends to drive u to u_R, since $g_R > 0$ and a is positive definite. This is a natural boundary condition for the heat conduction problem when the external world is held at a fixed temperature u_R and the body adjusts to it. We will no longer discuss the Robin condition, but instead concentrate on the Dirichlet and Neumann BC's.

The PDE (7.3) and the BC's (7.6)–(7.8) constitute our boundary value problem (BVP). As we will see, this problem is well posed, which means that there exists a unique solution to the system, and that it varies continuously in some norm with respect to changes in the data f, u_D, and g_N.

7.2. A Variational Problem and Minimization of Energy

For ease of exposition, let us consider the Dirichlet BVP

$$\begin{cases} -\nabla\cdot(a\nabla u) + cu = f & \text{in } \Omega\ , \\ \qquad\qquad\quad u = u_D & \text{on } \Gamma_D\ , \end{cases} \tag{7.9}$$

where we have set $b = 0$ and $\Gamma_D = \partial\Omega$. To make classical sense of this problem, we would expect $u \in C^2(\Omega) \cap C^0(\bar\Omega)$, so we would need to require that $f \in C^0(\Omega)$, $a \in (C^1(\Omega))^{d\times d}$, $c \in C^0(\Omega)$, and $u_D \in C^0(\partial\Omega)$. Often in practice these functions are not so well behaved, so we therefore interpret the problem in a weak or distributional sense.

If merely $f \in L_2(\Omega)$, $a \in (W^{1,\infty}(\Omega))^{d\times d}$, and $c \in L_\infty(\Omega)$, then we should expect $u \in H^2(\Omega)$. Moreover, then $u|_{\partial\Omega} \in H^{3/2}(\partial\Omega)$ is well defined by the trace theorem. Thus the BVP has a mathematically precise and consistent meaning formulated as: If f, a, and c are as stated and $u_D \in H^{3/2}(\partial\Omega)$, then find $u \in H^2(\Omega)$ such that (7.9) holds. This is *not* an easy problem; fortunately, we can find a better formulation using ideas of duality from distribution theory.

We first proceed formally: we will justify the calculations a bit later. We first multiply the PDE by a test function $v \in \mathcal{D}(\Omega)$, integrate in x, and integrate by parts. This is

$$\int_\Omega (-\nabla\cdot(a\nabla u) + cu)\, v\, dx = \int_\Omega (a\nabla u\cdot\nabla v + cuv)\, dx = \int_\Omega fv\, dx\ .$$

We have evened out the required smoothness of u and v, requiring only that each has a single derivative. Now if we only ask that $f \in H^{-1}(\Omega)$, $a \in (L_\infty(\Omega))^{d\times d}$, and $c \in L_\infty(\Omega)$, then we should expect that $u \in H^1(\Omega)$; moreover, we merely need $v \in H^1_0(\Omega)$. This is much less restrictive than asking for $u \in H^2(\Omega)$, so it should be easier to find such a solution satisfying the PDE. Moreover, $u|_{\partial\Omega} \in H^{1/2}(\partial\Omega)$ is still a nice function, and only requires $u_D \in H^{1/2}(\partial\Omega)$.

REMARK. Above we wanted to take cu in the same space as f, which was trivially achieved for $c \in L_\infty(\Omega)$. The Sobolev Imbedding Theorem allows us to do better. For example, suppose indeed that $u \in H^1(\Omega)$ and that we want $cu \in L_2(\Omega)$ (to avoid negative index spaces). Then in fact $u \in L_q(\Omega)$ for any finite $q \le 2d/(d-2)$ if $d \ge 2$ and $u \in C_B(\Omega) \subset L_\infty(\Omega)$ if $d = 1$. Thus we can take

$$c \in \begin{cases} L_2(\Omega) & \text{if } d = 1 \ , \\ L_{2+\epsilon}(\Omega) & \text{if } d = 2 \text{ for any } \epsilon > 0 \ , \\ L_d & \text{if } d \ge 3 \ , \end{cases}$$

and obtain $cu \in L_2(\Omega)$ as desired.

With this reduced regularity requirement on u ($u \in H^1(\Omega)$, not $H^2(\Omega)$), we can reformulate the problem rigorously as a *variational problem.* Our PDE (7.9) involves a *linear* operator

$$A \equiv -\nabla \cdot a\nabla + c : H^1(\Omega) \to H^{-1}(\Omega) \ ,$$

which we will transform into a *bilinear* operator

$$B : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R} \ .$$

Assume that $u \in H^1(\Omega)$ solves the PDE (we will show existence of a solution later), and take a test function $v \in H^1_0(\Omega)$. Then

$$\langle -\nabla \cdot (a\nabla u) + cu, v\rangle_{H^{-1},H^1_0} = \langle f, v\rangle_{H^{-1},H^1_0} \ .$$

Let $\{v_j\}_{j=1}^\infty \subset \mathcal{D}(\Omega)$ be a sequence converging to v in $H^1_0(\Omega)$. Then

$$\begin{aligned} \langle -\nabla \cdot (a\nabla u), v\rangle_{H^{-1},H^1_0} &= \lim_{j\to\infty} \langle -\nabla \cdot a\nabla u, v_j\rangle_{H^{-1},H^1_0} \\ &= \lim_{j\to\infty} \langle -\nabla \cdot a\nabla u, v_j\rangle_{\mathcal{D}',\mathcal{D}} \\ &= \lim_{j\to\infty} \langle a\nabla u, \nabla v_j\rangle_{\mathcal{D}',\mathcal{D}} \\ &= \lim_{j\to\infty} (a\nabla u, \nabla v_j)_{L_2(\Omega)} \\ &= (a\nabla u, \nabla v)_{L_2(\Omega)} \ , \end{aligned}$$

where the "$L_2(\Omega)$"-inner product is actually the one for $(L_2(\Omega))^d$. Thus

$$(a\nabla u, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)} = \langle f, v\rangle_{H^{-1},H^1_0} \ .$$

Let us define B by

$$B(u, v) = (a\nabla u, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)} \qquad \forall u, v \in H^1(\Omega) \ ,$$

and $F : H^1_0(\Omega) \to \mathbb{R}$ by

$$F(v) = \langle f, v\rangle_{H^{-1},H^1_0} \ ,$$

then the PDE has been reduced to the variational problem:

Find $u \in H^1(\Omega)$ such that

$$B(u, v) = F(v) \qquad \forall v \in H^1_0(\Omega) \ .$$

What about the boundary condition? Recall that the trace operator

$$\gamma_0 : H^1(\Omega) \xrightarrow{\text{onto}} H^{1/2}(\partial\Omega) \ .$$

Thus there is some $\tilde{u}_D \in H^1(\Omega)$ such that $\gamma_0(\tilde{u}_D) = u_D \in H^{1/2}(\partial\Omega)$. It is therefore required that

$$u \in H_0^1(\Omega) + \tilde{u}_D \ ,$$

so that $\gamma_0(u) = \gamma_0(\tilde{u}_D) = u_D$. For convenience, we no longer distinguish between u_D and its extension $\tilde{u}_D$. We summarize our construction below.

THEOREM 7.1. *If $\Omega \subset \mathbb{R}^d$ is a domain with a Lipschitz boundary, and $f \in H^{-1}(\Omega)$, $a \in (L_\infty(\Omega))^{d\times d}$, $c \in L_\infty(\Omega)$, and $u_D \in H^1(\Omega)$, then the BVP for $u \in H^1(\Omega)$,*

$$\begin{cases} -\nabla\cdot(a\nabla u) + cu = f & \text{in } \Omega \ , \\ \qquad\qquad\quad u = u_D & \text{on } \partial\Omega \ , \end{cases} \tag{7.10}$$

is equivalent to the variational problem:

Find $u \in H_0^1(\Omega) + u_D$ such that

$$B(u,v) = F(v) \qquad \forall v \in H_0^1(\Omega) \ , \tag{7.11}$$

where $B : H^1(\Omega)\times H^1(\Omega) \to \mathbb{R}$ is

$$B(u,v) = (a\nabla u, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)}$$

and $F : H_0^1(\Omega) \to \mathbb{R}$ is

$$F(v) = \langle f, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \ .$$

Actually, we showed that a solution to the BVP (7.10) gives a solution to the variational problem (7.11). By reversing the steps above, we see the converse implication. Note also that above we have extended the integration by parts formula (7.2) to the case where $\phi = v \in H_0^1(\Omega)$ and merely $\psi = -a\nabla u \in (L_2(\Omega))^d$.

The connection between the BVP (7.10) and the variational problem (7.11) is further illuminated by considering the following *energy functional.*

DEFINITION. If a symmetric (i.e., $a = a^T$), then the *energy functional* $J : H_0^1(\Omega) \to \mathbb{R}$ for (7.10) is given by

$$\begin{aligned} J(v) = \tfrac{1}{2}\big[&(a\nabla v,\nabla v)_{L_2(\Omega)} + (cv,v)_{L_2(\Omega)}\big] \\ &- \langle f, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} + (a\nabla u_D, \nabla v)_{L_2(\Omega)} + (cu_D, v)_{L_2(\Omega)} \ . \end{aligned} \tag{7.12}$$

We will study the calculus of variations in Chapter 8; however, we can easily make a simple computation here. We claim that any solution of (7.10), minus u_D, minimizes the "energy" $J(v)$. To see this, let $v \in H_0^1(\Omega)$ and compute

$$\begin{aligned} J(u - u_D + v) - J(u - u_D) &= (a\nabla u, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)} - \langle f, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \\ &\quad + \tfrac{1}{2}\big[(a\nabla v,\nabla v)_{L_2(\Omega)} + (cv,v)_{L_2(\Omega)}\big] \ , \end{aligned} \tag{7.13}$$

using that a is symmetric. If u satisfies (7.11), then

$$J(u - u_D + v) - J(u - u_D) = \tfrac{1}{2}\big[(a\nabla v,\nabla v)_{L_2(\Omega)} + (cv,v)_{L_2(\Omega)}\big] \ge 0 \ ,$$

provided that a is positive definite and $c \ge 0$. Thus every function in $H_0^1(\Omega)$ has "energy" at least as great as $u - u_D$.

Conversely, if $u - u_D \in H_0^1(\Omega)$ is to minimize the energy $J(v)$, then replacing in (7.13) v by ϵv for $\epsilon \in \mathbb{R}$, $\epsilon \neq 0$, we see that the difference quotient

$$\begin{aligned}\tfrac{1}{\epsilon}\big[J(u-u_D+\epsilon v) - J(u-u_D)\big] &= (a\nabla u, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)} - \langle f, v\rangle_{H^{-1}(\Omega),H_0^1(\Omega)} \\ &\quad + \tfrac{\epsilon}{2}\big[(a\nabla v,\nabla v)_{L_2(\Omega)} + (cv,v)_{L_2(\Omega)}\big] \ ,\end{aligned} \tag{7.14}$$

must be nonnegative if $\epsilon > 0$ and nonpositive if $\epsilon < 0$. Taking $\epsilon \to 0$ on the right-hand side shows that the first three terms must be both nonnegative and nonpositive, i.e., zero; thus, u must satisfy (7.11). Note that as $\epsilon \to 0$, the left-hand side is a kind of derivative of J at $u - u_D$. At the minimum, we have a critical point where the derivative vanishes.

THEOREM 7.2. *If the hypotheses of Theorem 7.1 hold, and if $c \geq 0$ and a is symmetric and positive definite, then* (7.10) *and* (7.11) *are also equivalent to the minimization problem:*

Find $u \in H_0^1(\Omega) + u_D$ such that

$$J(u-u_D) \leq J(v) \qquad \forall v \in H_0^1(\Omega) \ , \tag{7.15}$$

where J is given above by (7.12).

The physical principles of conservation or energy minimization are equivalent in this context, and they are connected by the variational problem: (1) it is the *weak form* of the BVP, given by multiplying by a test function, integrating, and integrating by parts to even out the number of derivatives on the solution and the test function, and (2) the variational problem also gives the critical point of the energy functional where it is minimized.

7.3. The Closed Range Theorem and operators bounded below

We continue with an abstract study of equation solvability that will be needed in the next section. In this section, we do not require the field to be real. We begin with a basic definition.

DEFINITION. Let X be a NLS and $Z \subset X$. Then the *orthogonal complement* of Z is

$$Z^\perp = \{x^* \in X^* : \langle x^*, z\rangle_{X^*,X} = 0 \ \forall z \in Z\} \ .$$

PROPOSITION 7.3. *Let X be a NLS and $Z \subset X$ a linear subspace. Then the following hold:*

(a) *$Z^\perp$ is closed in X^*;*
(b) *$Z \subset (Z^\perp)^\perp$;*
(c) *Z is closed in X if and only if $Z = (Z^\perp)^\perp$.*

The linearity of Z is needed only for (c). Of course, $(Z^\perp)^\perp \subset X^{**}$, so we have used the natural inclusion $X \subset X^{**}$ implicitly above.

PROOF. For (a), suppose that we have a sequence $\{y_j\}_{j=1}^\infty \subset Z^\perp$ that converges in X^* to y. But then for any $z \in Z$,

$$0 = \langle y_j, z\rangle_{X^*,X} \to \langle y, z\rangle_{X^*,X} \ ,$$

so $y \in Z^\perp$ and $Z^\perp$ is closed. Result (b) is a direct consequence of the definitions: for $z \in Z \subset X \subset X^{**}$ we want that $z \in (Z^\perp)^\perp$, i.e., that $\langle z, y\rangle_{X^{**},X^*} = \langle z, y\rangle_{X,X^*} = 0$ for all $y \in Z^\perp$, which holds.

Finally, for (c), that Z is closed follows from (a). For the other implication suppose Z is closed. We have (b), so we only need to show that $(Z^\perp)^\perp \subset Z$. Suppose that there is some nonzero $x \in (Z^\perp)^\perp$ such that $x \notin Z$. Now the Hahn-Banach Theorem, specifically Lemma 2.15, gives us the existence of $f \in ((Z^\perp)^\perp)^*$ such that $f(x) \neq 0$ but $f(z) = 0$ for all $z \in Z$, since Z is linear. That is, $f \in Z^\perp$, so x cannot be in $(Z^\perp)^\perp$, a contradiction. $\square$

Proposition 7.4. *Let X and Y be NLS's and $A : X \to Y$ a bounded linear operator. Then*

$$R(A)^{\perp} = N(A^*) \ ,$$

where $R(A)$ is the range of A and $N(A^*)$ is the null space of A^*.

Proof. We note that $y \in R(A)^{\perp}$ if and only if for every $x \in X$,

$$0 = \langle y, Ax \rangle_{Y^*,Y} = \langle A^* y, x \rangle_{X^*,X} \ ,$$

which is true if and only if $A^* y = 0$. $\square$

We have now immediately the following important theorem.

Theorem 7.5 (Closed Range Theorem). *Let X and Y be NLS's and $A : X \to Y$ a bounded linear operator. Then $R(A)$ is closed in Y if and only if $R(A) = N(A^*)^{\perp}$.*

This theorem has implications for a class of operators that often arise.

Definition. Let X and Y be NLS's and $A : X \to Y$. We say that A is *bounded below* if there is some constant $\gamma > 0$ such that

$$\|Ax\|_Y \geq \gamma \|x\|_X \qquad \forall x \in X \ .$$

A linear operator that is bounded below is one-to-one. If it also mapped onto Y, it would have a continuous inverse. We can determine whether $R(A) = Y$ by the Closed Range Theorem.

Theorem 7.6. *Let X and Y be Banach spaces and $A : X \to Y$ a continuous linear operator. Then the following are equivalent:*

(a) *A is bounded below;*
(b) *A is injective and $R(A)$ is closed;*
(c) *A is injective and $R(A) = N(A^*)^{\perp}$.*

Proof. The Closed Range Theorem gives the equivalence of (b) and (c). Suppose (a). Then A is injective. Let $\{y_j\}_{j=1}^{\infty} \subset R(A)$ converge to $y \in Y$. Choose $x_j \in X$ so that $Ax_j = y_j$ (the choice is unique), and note that

$$\|y_j - y_k\|_Y = \|A(x_j - x_k)\|_Y \geq \gamma \|x_j - x_k\|_X$$

implies that $\{x_j\}_{j=1}^{\infty}$ is Cauchy. Let $x_j \to x \in X$ and define $y = Ax \in R(A)$. Since A is continuous, $y_j = Ax_j \to Ax = y$, and $R(A)$ is closed.

Conversely, suppose (b). Then $R(A)$, being closed, is a Banach space itself. Thus $A : X \to R(A)$ is invertible, with continuous inverse by the Open Mapping Theorem 2.22. For $x \in X$, compute

$$\|x\|_X = \|A^{-1}Ax\|_X \leq \|A^{-1}\| \|Ax\|_Y \ ,$$

which gives (a) with constant $\gamma = 1/\|A^{-1}\|$. $\square$

Corollary 7.7. *Let X and Y be Banach spaces and $A : X \to Y$ a continuous linear operator. Then A is continuously invertible if and only if A is bounded below and $N(A^*) = \{0\}$ (i.e., A^* is injective).*

7.4. The Lax-Milgram Theorem

It is easy at this stage to prove existence of a unique solution to (7.10), or equivalently, (7.11), provided that a is symmetric and uniformly positive definite, $c \geq 0$, and both these functions are bounded. This is because $B(\cdot,\cdot)$ is then an inner-product on $H_0^1(\Omega)$, and this inner-product is equivalent to the usual one. To see these facts, we easily note that B is bilinear and symmetric (since a is symmetric), and $B(v,v) \geq 0$. We will show that $B(v,v) = 0$ implies $v = 0$ in a moment, which will show that B is an inner-product. For the equivalence with the $H_0^1(\Omega)$ inner-product, we have the upper bound

$$B(v,v) = (a\nabla v, \nabla v) + (cv, v) \leq \|a\|_{(L_\infty(\Omega))^{d\times d}} \|\nabla v\|_{L_2(\Omega)}^2 + \|c\|_{L_\infty(\Omega)} \|v\|_{L_2(\Omega)}^2 \leq C_1 \|v\|_{H_0^1(\Omega)}^2 ,$$

for some constant C_1. A lower bound is easy to obtain if c is *strictly* positive, i.e., bounded below by a positive constant. But we allow merely $c \geq 0$ by using the Poincaré inequality, which is a direct consequence of Cor. 6.18.

THEOREM 7.8 (Poincaré Inequality). *If $\Omega \subset \mathbb{R}^d$ is bounded, then there is some constant C such that*

$$\|v\|_{H_0^1(\Omega)} \leq C \|\nabla v\|_{L_2(\Omega)} \qquad \forall v \in H_0^1(\Omega) . \tag{7.16}$$

Now we have that

$$B(v,v) = (a\nabla v, \nabla v) + (cv, v) \geq a_* \|\nabla v\|_{L_2(\Omega)}^2 \geq (a_*/C^2) \|v\|_{H_0^1(\Omega)}^2 ,$$

and now both $B(v,v) = 0$ implies $v = 0$ and the equivalence of norms is established.

Problem (7.11) becomes:

Find $w = u - u_D \in H_0^1(\Omega)$ such that

$$B(w,v) = F(v) - B(u_D, v) \equiv \tilde{F}(v) \quad \forall v \in H_0^1(\Omega).$$

Now $\tilde{F} : H_0^1(\Omega) \to \mathbb{R}$ is linear and bounded:

$$|\tilde{F}(v)| \leq |F(v)| + |B(u_D, v)| \leq \big(\|F\|_{H^{-1}(\Omega)} + C\|u_D\|_{H^1(\Omega)}\big) \|v\|_{H_0^1(\Omega)} ,$$

where, again, C depends on the $L_\infty(\Omega)$-norms of a and c. Thus $\tilde{F} \in (H_0^1(\Omega))^* = H^{-1}(\Omega)$, and we seek to represent $\tilde{F}$ as $w \in H_0^1(\Omega)$ through the inner-product B. The Riesz Representation Theorem 3.12 gives us a unique such w. We have proved the following theorem.

THEOREM 7.9. *If $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain, $f \in H^{-1}(\Omega)$, $u_D \in H^1(\Omega)$, $a \in (L_\infty(\Omega))^{d\times d}$ is uniformly positive definite and symmetric on Ω, and $c \geq 0$ is in $L_\infty(\Omega)$, then there is a unique solution $u \in H^1(\Omega)$ to the BVP* (7.10) *and, equivalently, the variational problem* (7.11). *Moreover, there is a constant $C > 0$ such that*

$$\|u\|_{H^1(\Omega)} \leq C\big(\|F\|_{H^{-1}(\Omega)} + \|u_D\|_{H^1(\Omega)}\big) . \tag{7.17}$$

This last inequality is a consequence of the facts that

$$\|u\|_{H^1(\Omega)} \leq \|w\|_{H_0^1(\Omega)} + \|u_D\|_{H^1(\Omega)}$$

and

$$\|w\|_{H_0^1(\Omega)}^2 \leq C B(w,w) = C\tilde{F}(w) \leq C\big(\|F\|_{H^{-1}(\Omega)} + \|u_D\|_{H^1(\Omega)}\big) \|w\|_{H_0^1(\Omega)} .$$

REMARK. We leave it as an exercise to show that u is independent of the extension of u_D from $\partial\Omega$ to all of Ω. This extension is not unique, and we have merely that once the extension for u_D is fixed, then w is unique. That is, w depends on the extension. The reader should show that the sum $u = w + u_D$ does not depend on the extension chosen. Moreover, since the extension operator is bounded, that is,

$$\|u_D\|_{H^1(\Omega)} \le C\|u_D\|_{H^{1/2}(\partial\Omega)} ,$$

we can modify (7.17) so that it reads

$$\|u\|_{H^1(\Omega)} \le C\big(\|F\|_{H^{-1}(\Omega)} + \|u_D\|_{H^{1/2}(\partial\Omega)}\big) ,$$

and thereby refers only to the raw data itself and not the extension.

For more general problems, where either a is not symmetric, or $b \ne 0$ in the original Dirichlet problem (7.9), B is no longer symmetric, so it cannot be an inner-product. We need a generalization of the Riesz theorem to handle this case. In fact, we present this generalization for Banach spaces rather than restricting to Hilbert spaces.

THEOREM 7.10 (Generalized Lax-Milgram Theorem). *Let $\mathcal{X}$ and Y be real Banach spaces, and suppose that Y is reflexive, $B : \mathcal{X} \times Y \to \mathbb{R}$ is bilinear, and $X \subset \mathcal{X}$ be a closed subspace. Assume also the following three conditions:*

(a) *B is continuous on $\mathcal{X} \times Y$, i.e., there is some $M > 0$ such that*

$$|B(x,y)| \le M\|x\|_{\mathcal{X}}\|y\|_Y \qquad \forall x \in \mathcal{X},\ y \in Y ;$$

(b) *B satisfies the* inf-sup *condition on $X \times Y$, i.e., there is some $\gamma > 0$ such that*

$$\inf_{\substack{x\in X\\ \|x\|_{\mathcal{X}}=1}} \sup_{\substack{y\in Y\\ \|y\|_Y=1}} B(x,y) \ge \gamma > 0 ;$$

(c) *and B satisfies the nondegeneracy condition on X that*

$$\sup_{x\in X} B(x,y) > 0 \qquad \forall y \in Y,\ y \ne 0 .$$

If $x_0 \in \mathcal{X}$ and $F \in Y^*$, then there is a unique u solving the abstract variational problem:

Find $u \in X + x_0 \subset \mathcal{X}$ such that

$$B(u,v) = F(v) \qquad \forall v \in Y . \tag{7.18}$$

Moreover,

$$\|u\|_{\mathcal{X}} \le \frac{1}{\gamma}\|F\|_{Y^*} + \left(\frac{M}{\gamma} + 1\right)\|x_0\|_{\mathcal{X}} . \tag{7.19}$$

We remark that (b) is often written equivalently as

$$\sup_{\substack{y\in Y\\ y\ne 0}} \frac{B(x,y)}{\|y\|_Y} \ge \gamma\|x\|_{\mathcal{X}} \qquad \forall x \in X .$$

In our context, $\mathcal{X} = H^1(\Omega)$, $X = Y = H^1_0(\Omega)$, and $x_0 = u_D$.

PROOF. Assume first that $x_0 = 0$. For each fixed $x \in X$, $B(x,\cdot)$ defines a linear functional on Y, since B is linear in each variable separately, so certainly the second. Let A represent the operator that takes x to $B(x,y)$:

$$\langle Ax, y\rangle = Ax(y) \equiv B(x,y) \qquad \forall x \in X,\ y \in Y .$$

Since (a) gives that

$$|\langle Ax, y\rangle| = |B(x,y)| \le (M\|x\|_{\mathcal{X}})\|y\|_Y \ ,$$

Ax is a continuous linear functional, i.e., $A : X \to Y^*$. Moreover, A itself is linear, since B is linear in its first variable, and therefore A is a continuous linear operator:

$$\|Ax\|_{Y^*} = \sup_{\|y\|_Y=1} \langle Ax, y\rangle \le M\|x\|_{\mathcal{X}} \ .$$

We reformulate (7.18) in terms of A as the problem of finding $u \in X$ such that

$$Au = F \ .$$

Now (b) implies that

$$\|Ax\|_{Y^*} \ge \gamma\|x\|_{\mathcal{X}} \qquad \forall x \in X \ , \tag{7.20}$$

so A is bounded below and u, if it exists, must be unique (i.e., A is one-to-one). Since X is closed (Theorem 7.6), it is a Banach space and we conclude that the range of A, $R(A)$, is closed in Y^*. The Closed Range Theorem 7.5 now implies that $R(A) = N(A^*)^\perp$. We wish to show that $N(A^*) = \{0\}$, so that A maps onto. Suppose that for some $y \in Y = Y^{**}$, $y \in N(A^*)$; that is,

$$B(x,y) = \langle Ax, y\rangle = 0 \qquad \forall x \in X \ .$$

But (c) implies then that $y = 0$. So we have that A has a bounded inverse, with $\|A^{-1}\| \le 1/\gamma$ by (7.20), and $u = A^{-1}F$ solves our problem.

Finally, we compute

$$\|u\|_{\mathcal{X}} = \|A^{-1}F\|_{\mathcal{X}} \le \|A^{-1}\|\|F\|_{Y^*} \le \frac{1}{\gamma}\|F\|_{Y^*} \ .$$

The theorem is established when $x_0 = 0$.

If $x_0 \ne 0$, we reduce to the previous case, since (7.18) is equivalent to:

Find $w \in X$ such that

$$B(w,v) = \tilde{F}(v) \qquad \forall v \in Y \ ,$$

where $u = w + x_0 \in X + x_0 \subset \mathcal{X}$ and

$$\tilde{F}(v) = F(v) - B(x_0, v) \ .$$

Now $\tilde{F} \in Y^*$ and

$$|\tilde{F}(v)| \le |F(v)| + |B(x_0,v)| \le \big(\|F\|_{Y^*} + M\|x_0\|_{\mathcal{X}}\big)\|v\|_Y \ .$$

Thus the previous result gives

$$\|w\|_{\mathcal{X}} \le \frac{1}{\gamma}\big(\|F\|_{Y^*} + M\|x_0\|_{\mathcal{X}}\big) \ ,$$

and so

$$\|u\|_{\mathcal{X}} \le \|w + x_0\|_{\mathcal{X}} \le \|w\|_{\mathcal{X}} + \|x_0\|_{\mathcal{X}}$$

gives the desired bound. □

When $X = Y$ is a Hilbert space, things are a bit simpler.

Corollary 7.11 (Lax-Milgram Theorem). *Let $\mathcal{X}$ be a real Hilbert space with closed subspace X. Let $B : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a bilinear functional satisfying the following two conditions:*

(i) *B is continuous on $\mathcal{X}$, i.e., there is some $M > 0$ such that*

$$|B(x,y)| \le M\|x\|_{\mathcal{X}}\|y\|_{\mathcal{X}} \qquad \forall x,y \in \mathcal{X} \ ;$$

(ii) *B is* coercive *(or* elliptic*) on X i.e., there is some $\gamma > 0$ such that*

$$B(x,x) \geq \gamma \|x\|_{\mathcal{X}}^2 \qquad \forall x \in X \ .$$

If $x_0 \in \mathcal{X}$ and $F \in X^*$, then there is a unique u solving the abstract variational problem:

Find $u \in X + x_0 \subset \mathcal{X}$ such that

$$B(u,v) = F(v) \qquad \forall v \in X \ . \tag{7.21}$$

Moreover,

$$\|u\|_{\mathcal{X}} \leq \frac{1}{\gamma}\|F\|_{X^*} + \left(\frac{M}{\gamma} + 1\right)\|x_0\|_{\mathcal{X}} \ . \tag{7.22}$$

Proof. The corollary is just a special case of the theorem except that (ii) has replaced (b) and (c). We claim that (ii) implies both (b) and (c), so the corollary follows.

Easily, we have (c), since for any $y \in X$,

$$\sup_{x\in X} B(x,y) \geq B(y,y) \geq \gamma \|y\|_{\mathcal{X}}^2 > 0$$

whenever $y \neq 0$. Similarly, for any $x \in X$ with norm one,

$$\sup_{\substack{y\in X \\ \|y\|_{\mathcal{X}}=1}} B(x,y) \geq B(x,x) \geq \gamma > 0 \ ,$$

so the infimum over all such x is bounded below by γ, which is (b). $\square$

The Generalized Lax-Milgram Theorem gives the existence of a bounded linear solution operator $S : Y^* \times \mathcal{X} \to \mathcal{X}$ such that $S(F, x_0) = u \in X + x_0 \subset \mathcal{X}$ satisfies

$$B(S(F,x_0),v) = F(v) \qquad \forall v \in Y \ .$$

The bound on S is given by (7.19). This bound shows that the solution varies continuously with the data. That is, by linearity,

$$\|S(F,x_0) - S(G,y_0)\|_{\mathcal{X}} \leq \frac{1}{\gamma}\|F - G\|_{X^*} + \left(\frac{M}{\gamma} + 1\right)\|x_0 - y_0\|_{\mathcal{X}} \ .$$

So if the data (F, x_0) is perturbed a bit to (G, y_0), then the solution $S(F, x_0)$ changes by a small amount to $S(G, y_0)$, where the magnitudes of the changes are measured in the norms as above.

7.5. Application to second order elliptic equations

We consider again the BVP (7.10), in the form of the variational problem (7.11). To apply the Lax-Milgram Theorem, we set $\mathcal{X} = H^1(\Omega)$, $X = Y = H_0^1(\Omega)$, and $x_0 = u_D$. Now $B : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is continuous, since a and c are bounded:

$$\begin{aligned} |B(u,v)| &= |(a\nabla u, \nabla v)_{L_2(\Omega)} + (cu,v)_{L_2(\Omega)}| \\ &\leq \|a\|_{(L_\infty(\Omega))^{d\times d}}\|\nabla u\|_{L_2(\Omega)}\|\nabla v\|_{L_2(\Omega)} + \|c\|_{L_\infty(\Omega)}\|u\|_{L_2(\Omega)}\|v\|_{L_2(\Omega)} \\ &\leq M\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)} \ , \end{aligned}$$

by Hölder's inequality for some $M > 0$ depending on the bounds for a and c. Coercivity is more interesting. We will only assume that $c \geq 0$, since in practice, often $c = 0$. Using that a is uniformly positive definite and Ω is bounded, we compute

$$\begin{aligned}B(u,u) &= (a\nabla u, \nabla u)_{L_2(\Omega)} + (cu,u)_{L_2(\Omega)}\\ &\geq a_*(\nabla u,\nabla u)_{L_2(\Omega)} = a_*\|\nabla u\|^2_{L_2(\Omega)}\\ &\geq (a_*/C^2)\|u\|^2_{H^1(\Omega)}\ ,\end{aligned}$$

for some $C > 0$ by Poincaré's inequality. Thus there exists a unique solution $u \in H^1_0(\Omega) + u_D$, and

$$\|u\|_{H^1(\Omega)} \leq \frac{C^2}{a_*}\|f\|_{H^{-1}(\Omega)} + \left(\frac{C^2 M}{a_*} + 1\right)\|u_D\|_{H^1(\Omega)}\ .$$

Note that the boundary condition $u = u_D$ on $\partial\Omega$ is enforced by out selection of the *trial space* $H^1_0(\Omega) + u_D$, i.e., the space within which we seek a solution has every member satisfying the boundary condition. Because of this, we call the Dirichlet BC an *essential* BC for this problem.

7.5.1. The general Dirichlet problem. Consider more generally the full elliptic equation (7.3) with a Dirichlet BC:

$$\begin{cases} -\nabla\cdot(a\nabla u + bu) + cu = f & \text{in } \Omega\ ,\\ \hfill u = u_D & \text{on } \partial\Omega\ .\end{cases}$$

We leave it to the reader to show that an equivalent variational problem is:

Find $u \in H^1_0(\Omega) + u_D$ such that

$$B(u,v) = F(v) \qquad \forall v \in H^1_0(\Omega)\ ,$$

where

$$\begin{aligned}B(u,v) &= (a\nabla u,\nabla v)_{L_2(\Omega)} + (bu,\nabla v)_{L_2(\Omega)} + (cu,v)_{L_2(\Omega)}\ ,\\ F(v) &= \langle f,v\rangle_{H^{-1}(\Omega),H^1_0(\Omega)}\ .\end{aligned}$$

Now if $b \in (L_\infty(\Omega))^d$ (and a and c are bounded as before), then the bilinear form is bounded. For coercivity, assume again that $c \geq 0$ and a is uniformly positive definite. Then for $v \in H^1_0(\Omega)$,

$$\begin{aligned}B(v,v) &= (a\nabla v,\nabla v)_{L_2(\Omega)} + (bv,\nabla v)_{L_2(\Omega)} + (cv,v)_{L_2(\Omega)}\\ &\geq a_*\|\nabla v\|^2_{L_2(\Omega)} - |(bv,\nabla v)_{L_2(\Omega)}|\\ &\geq \big(a_*\|\nabla v\|_{L_2(\Omega)} - \|b\|_{(L_\infty(\Omega))^d}\|v\|_{L_2(\Omega)}\big)\|\nabla v\|_{L_2(\Omega)}\ .\end{aligned}$$

Poincaré's inequality tells us that for some $C_P > 0$,

$$a_*\|\nabla v\|_{L_2(\Omega)} - \|b\|_{(L_\infty(\Omega))^d}\|v\|_{L_2(\Omega)} \geq \big(a_* - C_P\|b\|_{(L_\infty(\Omega))^d}\big)\|\nabla v\|_{L_2(\Omega)}\ .$$

To continue in the present context, we must assume that for some $\alpha > 0$,

$$a_* - C_P\|b\|_{(L_\infty(\Omega))^d} \geq \alpha > 0\ ; \tag{7.23}$$

this restricts the size of b relative to a. Then we have that

$$B(v,v) \geq \alpha\|\nabla v\|^2_{L_2(\Omega)} \geq \frac{\alpha}{C_P^2+1}\|v\|^2_{H^1(\Omega)}\ ,$$

and the Lax-Milgram Theorem gives us a unique solution to the problem as well as the continuous dependence result. Note that in this general case, if a is not symmetric or $b \neq 0$, then B is not symmetric, so B cannot be an inner-product. However, continuity and coercivity show that the diagonal of B (i.e., $u = v$) is equivalent to the square of the $H^1_0(\Omega)$-norm.

7.5.2. The Neumann problem with lowest order term. We turn now to the Neumann BVP

$$\begin{cases} -\nabla\cdot(a\nabla u)+cu=f & \text{in } \Omega\ , \\ \qquad\quad -a\nabla u\cdot\nu=g & \text{on } \partial\Omega\ , \end{cases} \tag{7.24}$$

wherein we have set $b=0$ for simplicity. This problem is more delicate than the Dirichlet problem, since for $u\in H^1(\Omega)$, we have no meaning in general for $a\nabla u\cdot\nu$. We proceed formally to derive a variational problem by assuming that u and the test function v are in, say $C^\infty(\bar\Omega)$. Then the Divergence Theorem can be applied to obtain

$$-\int_\Omega \nabla\cdot(a\nabla u)\,v\,dx=\int_\Omega a\nabla u\cdot\nabla v\,dx-\int_{\partial\Omega} a\nabla u\cdot\nu\,v\,dx\ ,$$

or, using the boundary condition and assuming that f and g are nice functions,

$$(a\nabla u,\nabla v)_{L_2(\Omega)}+(cu,v)_{L_2(\Omega)}=(f,v)_{L_2(\Omega)}-(g,v)_{L_2(\partial\Omega)}\ .$$

These integrals are well defined on $H^1(\Omega)$, so we have the variational problem:

Find $u\in H^1(\Omega)$ such that

$$B(u,v)=F(v)\qquad \forall v\in H^1(\Omega)\ , \tag{7.25}$$

where $B:H^1(\Omega)\times H^1(\Omega)\to\mathbb{R}$ is

$$B(u,v)=(a\nabla u,\nabla v)_{L_2(\Omega)}+(cu,v)_{L_2(\Omega)}$$

and $F:H^1(\Omega)\to\mathbb{R}$ is

$$F(v)=\langle f,v\rangle_{(H^1(\Omega))^*,H^1(\Omega)}-\langle g,v\rangle_{H^{-1/2}(\Omega),H^{1/2}(\Omega)}\ . \tag{7.26}$$

It is clear that we will require that $f\in(H^1(\Omega))^*$. Moreover, for $v\in H^1(\Omega)$, its trace is in $H^{1/2}(\Omega)$, so we merely require $g\in H^{-1/2}(\Omega)$, the dual of $H^{1/2}(\Omega)$.

A solution of (7.25) will be called a *weak solution* of (7.24). These problems are not strictly equivalent, because of the boundary condition. For the PDE, consider u satisfying the variational problem. Restrict to test functions $v\in\mathcal{D}(\Omega)$ to avoid $\partial\Omega$ and use the Divergence Theorem, as in the case of the Dirichlet boundary condition, to see that the differential equation in (7.24) is satisfied in the sense of distributions. This argument can be reversed to see that a solution in $H^1(\Omega)$ to the PDE gives a solution to the variational problem for $v\in\mathcal{D}(\Omega)$, and for $v\in H^1_0(\Omega)$ by density. The boundary condition will be satisfied only in some weak sense, i.e., only in the sense of the variational form.

If in fact the solution happens to be in, say, $H^2(\Omega)$, then $a\nabla u\cdot\nu\in H^{1/2}(\partial\Omega)$ and the argument above can be modified to show that indeed $-a\nabla u\cdot\nu=g$. Of course in this case, we must then have that $g\in H^{1/2}(\partial\Omega)$, and, moreover, that $f\in L_2(\Omega)$. So suppose that $u\in H^2(\Omega)$ solves the variational problem (and f and g are as stated). Restrict now to test functions $v\in H^1(\Omega)\cap C^\infty(\bar\Omega)$ to show that

$$\begin{aligned} B(u,v)&=(a\nabla u,\nabla v)_{L_2(\Omega)}+(cu,v)_{L_2(\Omega)}\\ &=-(\nabla\cdot(a\nabla u),v)_{L_2(\Omega)}+(a\nabla u\cdot\nu,v)_{L_2(\partial\Omega)}+(cu,v)_{L_2(\Omega)}\\ &=F(v)=(f,v)_{L_2(\Omega)}-(g,v)_{L_2(\partial\Omega)}\ . \end{aligned}$$

Using test functions $v\in C_0^\infty$ shows again by the Lebesgue Lemma that the PDE is satisfied. Thus, we have that

$$(a\nabla u\cdot\nu,v)_{L_2(\partial\Omega)}=-(g,v)_{L_2(\partial\Omega)}\ ,$$

and another application of the Lebesgue Lemma (this time on $\partial\Omega$) shows that indeed $-a\nabla u\cdot\nu = g$ in $L_2(\partial\Omega)$, and therefore also in $H^{1/2}(\partial\Omega)$. That is, a smoother solution of (7.25) also solves (7.24). The converse can be shown to hold as well by reversing the steps above, up to the statement that indeed $u \in H^2(\Omega)$. But this latter fact follows from the Elliptic Regularity Theorem 7.13 to be given at the end of this section.

Let us now apply the Lax-Milgram Theorem to our variational problem (7.25) to obtain the existence and uniqueness of a solution. We have seen that the bilinear form B is continuous if a and c are bounded functions. For coercivity, we require that a be uniformly positive definite and that c is uniformly positive: there exists $c_* > 0$ such that

$$c(x) \geq c_* > 0 \qquad \text{for a.e. } x \in \Omega\ .$$

This is required rather than merely $c \geq 0$ since $H^1(\Omega)$ does *not* satisfy a Poincaré inequality. Now we compute

$$\begin{aligned}(a\nabla u, \nabla u)_{L_2(\Omega)} + (cu, u)_{L_2(\Omega)} &\geq a_*\|\nabla u\|^2_{L_2(\Omega)} + c_*\|u\|^2_{L_2(\Omega)}\\ &\geq \min(a_*, c_*)\|u\|^2_{H^1(\Omega)}\ ,\end{aligned}$$

which is the coercivity of the form B. We now conclude that there is a unique solution of the variational problem (7.25) which varies continuously with the data. Moreover, if the solution is more regular (i.e., $u \in H^2(\Omega)$), then (7.24) has a solution as well. (But is it unique?)

Note that the boundary condition $-a\nabla u\cdot\nu = g$ on $\partial\Omega$ is *not* enforced by the *trial space* $H^1(\Omega)$, since most elements of this space do *not* satisfy the boundary condition. Rather, the BC is imposed in a weak sense as noted above. In this case, the Neumann BC is said to be a *natural* BC.

7.5.3. The Neumann problem with no zeroth order term. In this subsection, we also require that Ω be connected. If it is not, consider each connected piece separately.

Often the Neumann problem (7.24) is posed with $c \equiv 0$, in which case the problem is *degenerate* in the sense that coercivity of B is lost. In that case, the solution cannot be unique, since any constant function solves the homogeneous problem (i.e., the problem for data $f = g = 0$).

The problem is that the kernel of the operator ∇ is larger than $\{0\}$, and this kernel intersects the kernel of the boundary operator $-a\partial/\partial\nu$. In fact, this intersection is

$$Z = \{v \in H^1(\Omega) : v \text{ is constant a.e. on } \Omega\}\ ,$$

which is a closed subspace isomorphic to $\mathbb{R}$. If we "mod out" by $\mathbb{R}$, we can recover uniqueness. One way to do this is to insist that the solution have average zero. Let

$$\tilde{H}^1(\Omega) = \left\{u \in H^1(\Omega) : \int_\Omega u(x)\,dx = 0\right\},$$

which is isomorphic to $H^1(\Omega)/\mathbb{R}$, i.e., $H^1(\Omega)$ modulo constant functions, and so is a Hilbert space. To prove coercivity of B on $\tilde{H}^1(\Omega)$, we need a Poincaré inequality, which follows.

THEOREM 7.12. *If $\Omega \subset \mathbb{R}^d$ is a bounded and connected domain, then there is some constant $C > 0$ such that*

$$\|v\|_{L_2(\Omega)} \leq C\|\nabla v\|_{L_2(\Omega)} \qquad \forall v \in \tilde{H}^1(\Omega)\ . \tag{7.27}$$

PROOF. Suppose not. Then we can find a sequence $\{u_n\}_{n=1}^\infty \subset \tilde{H}^1(\Omega)$ such that

$$\|u_n\|_{L_2(\Omega)} = 1 \qquad \text{and} \qquad \|\nabla u_n\|_{L_2(\Omega)} < 1/n\ ,$$

and so

$$\nabla u_n \to 0 \qquad \text{strongly in } L_2(\Omega)\ .$$

Furthermore, $\|u\|_{H^1(\Omega)} \le \sqrt{2}$, so we conclude both that

$$u_n \xrightarrow{w} u \qquad \text{weakly in } H^1(\Omega)$$

and, by the Rellich-Kondrachov Theorem 6.20, for a subsequence,

$$u_n \to u \qquad \text{strongly in } L_2(\Omega)\ .$$

That is, $\nabla u_n \to 0$ and $\nabla u_n \xrightarrow{w} \nabla u$, so we conclude that $\nabla u = 0$. Thus u is a constant (since Ω is connected) and has average zero, so $u = 0$. But this contradicts the fact that

$$1 = \|u_n\|_{L_2(\Omega)} \to \|u\|_{L_2(\Omega)} = 0\ ,$$

and the inequality claimed in the theorem must hold. $\square$

On a connected domain, then, we have for $u \in \tilde{H}^1(\Omega)$

$$B(u,u) = (a\nabla u, \nabla v)_{L_2(\Omega)} \ge a_* \|\nabla u\|^2_{L_2(\Omega)} \ge C\|u\|^2_{H^1(\Omega)}$$

for some constant $C > 0$, that is, coercivity of B. Thus we conclude from the Lax-Milgram Theorem that a solution exists and is unique for the variational problem:

Find $u \in \tilde{H}^1(\Omega)$ such that

$$B(u,v) = F(v) \qquad \forall v \in \tilde{H}^1(\Omega)\ , \tag{7.28}$$

where $B(u,u) = (a\nabla u, \nabla v)_{L_2(\Omega)}$ and F is defined in (7.26), provided that $F \in (\tilde{H}^1(\Omega))^*$.

Often we prefer to formulate the Neumann problem in $H^1(\Omega)$ rather than in $\tilde{H}^1(\Omega)$ and accept the nonuniqueness. Actually, we pose the problem uniquely as:

Find $u \in \tilde{H}^1(\Omega)$ such that

$$B(u,v) = F(v) \qquad \forall v \in H^1(\Omega)\ . \tag{7.29}$$

In that case, for any $\alpha \in \mathbb{R}$,

$$B(u, v+\alpha) = B(u,v)\ ,$$

so if we have a solution $u \in H^1(\Omega)$, then also

$$F(v) = B(u,v) = B(u,v+\alpha) = F(v+\alpha) = F(v) + F(\alpha)$$

implies that $F(\alpha) = 0$ is required. That is, $\mathbb{R} \subset \ker(F)$. This condition is called a *compatibility condition,* and it says that the kernel of $B(u,\cdot)$ is contained in the kernel of F; that is, f and g must satisfy

$$\langle f, 1\rangle_{(H^1(\Omega))^*, H^1(\Omega)} - \langle g, 1\rangle_{H^{-1/2}(\Omega), H^{1/2}(\Omega)} = 0\ ,$$

which is to say

$$\int_\Omega f(x)\,dx = \int_{\partial\Omega} g(x)\,d\sigma(x)\ ,$$

provided that f and g are integrable.

The compatibility condition is necessary for obtaining a solution, but is it sufficient? Note that (c) of the Generalized Lax-Milgram Theorem 7.10 is *not* satisfied. The reformulated problem (7.29) only requires $F \in (H^1(\Omega))^*$. However, the compatibility condition actually says that $F \in (\tilde{H}^1(\Omega))^*$; moreover, it also says that we can restrict the test functions to $v \in \tilde{H}^1(\Omega)$. Thus (7.29) is equivalent to (7.28), which we already saw had a unique solution.

In abstract terms, we have the following situation. The problem is naturally posed for u and v in a Hilbert space X. However, there is nonuniqueness because the set $\{u \in X : B(u,v) = 0\ \forall v \in X\} = \{v \in X : B(u,v) = 0\ \forall u \in X\}$ is contained in the kernel of the natural BC. But the problem is well behaved when posed over X/Y, which requires $F \in (X/Y)^*$. The compatibility condition is precisely the condition that an element $F \in X^*$ is actually in $(X/Y)^*$.

7.5.4. Elliptic regularity. We close this section with an important result from the theory of elliptic PDE's. See, e.g., [**GT**] or [**Fo**] for a proof. This result can be used to prove the equivalence of the BVP and the variational problem in the case of Neumann BC's.

THEOREM 7.13 (Elliptic Regularity). *Suppose that $k \geq 0$ is an integer, $\Omega \subset \mathbb{R}^d$ is a bounded domain with a $C^{k+1,1}(\Omega)$-boundary, $a \in (W^{k+1,\infty}(\Omega))^{d\times d}$ is uniformly positive definite, $b \in (W^{k+1,\infty}(\Omega))^d$, and $c \in W^{k+2,\infty}(\Omega)$ is nonnegative. Suppose also that the bilinear form $B : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$,*

$$B(u,v) = (a\nabla u, \nabla v)_{L_2(\Omega)} + (bu, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)} \ ,$$

is continuous and coercive on X, for X given below.

(a) *If $f \in H^k(\Omega)$, $u_D \in H^{k+2}(\Omega)$, and $X = H^1_0(\Omega)$, then the Dirichlet problem:*
Find $u \in H^1_0(\Omega) + u_D$ such that

$$B(u,v) = (f,v)_{L_2(\Omega)} \qquad \forall v \in H^1_0(\Omega) \ , \tag{7.30}$$

has a unique solution $u \in H^{k+2}(\Omega)$ satisfying, for constant $C > 0$ independent of f, u, and u_D,

$$\|u\|_{H^{k+2}(\Omega)} \leq C\big(\|f\|_{H^k(\Omega)} + \|u_D\|_{H^{k+3/2}(\partial\Omega)}\big) \ .$$

Moreover, $k = -1$ is allowed in this case.

(b) *If $f \in H^k(\Omega)$, $g \in H^{k+1/2}(\partial\Omega)$, and $X = H^1(\Omega)$, then the Neumann problem:*
Find $u \in H^1(\Omega)$ such that

$$B(u,v) = (f,v)_{L_2(\Omega)} - (g,v)_{L_2(\partial\Omega)} \qquad \forall v \in H^1(\Omega) \ , \tag{7.31}$$

has a unique solution $u \in H^{k+2}(\Omega)$ satisfying, for constant $C > 0$ independent of f, u, and g,

$$\|u\|_{H^{k+2}(\Omega)} \leq C\big(\|f\|_{H^k(\Omega)} + \|g\|_{H^{k+1/2}(\partial\Omega)}\big) \ .$$

7.6. Galerkin approximations

Often we wish to find some simple approximation to our BVP. This could be for computational purposes, to obtain an explicit approximation of the solution, or for theoretical purposes to prove some property of the solution. We present here *Galerkin* methods, which give a framework for such approximation.

THEOREM 7.14. *Suppose that H is a Hilbert space with closed subspaces*

$$H_0 \subset H_1 \subset \cdots \subset H$$

such that the closure of $\displaystyle\bigcup_{n=0}^{\infty} H_n$ is H. Suppose also that $B : H \times H \to \mathbb{R}$ is a continuous, coercive bilinear form on H and that $F \in H^*$. Then the variational problems, one for each n,

Find $u_n \in H_n$ such that

$$B(u_n, v_n) = F(v_n) \qquad \forall v_n \in H_n \ , \tag{7.32}$$

have unique solutions. The same problem posed on H also has a unique solution $u \in H$, and

$$u_n \to u \qquad \text{in } H \ .$$

Moreover, if M and γ are respectively the continuity and coercivity constants for B, then for any n,

$$\|u - u_n\|_H \le \frac{M}{\gamma} \inf_{v_n \in H_n} \|u - v_n\|_H \ . \tag{7.33}$$

Furthermore, if B is symmetric, then for any n,

$$\|u - u_n\|_B \le \inf_{v_n \in H_n} \|u - v_n\|_B \ , \tag{7.34}$$

where $\|\cdot\|_B = B(\cdot,\cdot)^{1/2}$ is the energy norm.

Remark. Estimate (7.33) says that the approximation of u by u_n in H_n is *quasi-optimal* in the H-norm; that is, up to the constant factor M/γ, u_n is the best approximation to u in H_n. When B is symmetric, $\|\cdot\|_B$ is indeed a norm, as the reader can verify, equivalent to the H-norm by continuity and coercivity. Estimate (7.34) says that the Galerkin approximation $u_n \in H_n$ is optimal in the energy norm.

Proof. We have both

$$B(u_n, v_n) = F(v_n) \qquad \forall v_n \in H_n \ ,$$

and

$$B(u, v) = F(v) \qquad \forall v \in H \ .$$

Existence of unique solutions is given by the Lax-Milgram Theorem. Since $H_n \subset H$, restrict $v = v_n \in H_n$ in the latter and subtract to obtain that

$$B(u - u_n, v_n) = 0 \qquad \forall v_n \in H_n \ .$$

(We remark that in some cases B gives an inner-product, so in that case this relation says that the error $u - u_n$ is B-orthogonal to H_n; thus, this relation is referred to as *Galerkin orthogonality.*) Replace v_n by $(u - u_n) - (u - v_n) \in H_n$ for any $v_n \in H_n$ to obtain that

$$B(u - u_n, u - u_n) = B(u - u_n, u - v_n) \qquad \forall v_n \in H_n \ . \tag{7.35}$$

Thus,

$$\gamma \|u - u_n\|_H^2 \le B(u - u_n, u - u_n) = B(u - u_n, u - v_n) \le M \|u - u_n\|_H \|u - v_n\|_H \ ,$$

and (7.33) follows. If B is symmetric, then B is an inner-product, and the Cauchy-Schwarz inequality applied to (7.35) gives

$$\|u - u_n\|_B^2 = B(u - u_n, u - u_n) = B(u - u_n, u - v_n) \le \|u - u_n\|_B \|u - v_n\|_B \ ,$$

and (7.34) follows.

Finally, since $\bigcup_{n=0}^{\infty} H_n$ is dense in H, there are $\phi_n \in H_n$ such that $\phi_n \to u$ in H as $n \to \infty$. Then

$$\|u - u_n\|_H \le \frac{M}{\gamma} \inf_{v_n \in H_n} \|u - v_n\|_H \le \frac{M}{\gamma} \|u - \phi_n\|_H \ ,$$

so $u_n \to u$ in H as $n \to \infty$. $\square$

If (7.32) represents the equation for the critical point of an energy functional $J : H \to \mathbb{R}$, then for any n,

$$\inf_{v_n \in H_n} J(v_n) = J(u_n) \geq J(u) = \inf_{v \in H} J(v) \ .$$

That is, we find the function with minimal energy in the space H_n to approximate u. In this minimization form, the method is called a *Ritz* method.

In the theory of finite element methods, one attempts to define explicitly the spaces $H_n \subset H$ in such a way that the equations (7.32) can be solved easily and so that the optimal error

$$\inf_{v_n \in H_n} \|u - v_n\|_H$$

is quantifiably small. Such Galerkin finite element methods are extremely effective for computing approximate solutions to elliptic BVP's, and for many other types of equations as well. We now present a simple example.

EXAMPLE. Suppose that $\Omega = (0,1) \subset \mathbb{R}$ and $f \in L_2(0,1)$. Consider the BVP

$$\begin{cases} -u'' = f & \text{on } (0,1) \ , \\ u(0) = u(1) = 0 \ . \end{cases} \tag{7.36}$$

The equivalent variational problem is:

Find $u \in H_0^1(0,1)$ such that

$$(u', v')_{L_2} = (f, v)_{L_2} \qquad \forall v \in H_0^1(0,1) \ . \tag{7.37}$$

We now construct a suitable finite element decomposition of $H_0^1(0,1)$. Let $n \geq 1$ be an integer, and define $h = h_n = 1/n$ and a grid $x_i = ih$ for $i = 0, 1, ..., n$ of spacing h. Let

$$H_n = H_h = \{v \in C^0(0,1) : v(0) = v(1) = 0 \text{ and } v(x) \text{ is a first degree} \\ \text{polynomial on } [x_{i-1}, x_i] \text{ for } i = 1, 2, ..., n\} \ ;$$

that is, H_h consists of the continuous, piecewise linear functions. Note that $H_h \subset H_0^1(0,1)$, and H_h is a finite dimensional vector space. We leave it to the reader to show that the closure of $\displaystyle\bigcup_{n=1}^{\infty} H_h$ is dense in $H_0^1(0,1)$. In fact, one can show that there is a constant $C > 0$ such that for any $v \in H_0^1(0,1) \cap H^2(0,1)$,

$$\min_{v_h \in H_h} \|v - v_h\|_{H^1} \leq C \|v\|_{H^2} h \ . \tag{7.38}$$

The Galerkin finite element approximation is:

Find $u_h \in H_h$ such that

$$(u_h', v_h')_{L_2} = (f, v_h)_{L_2} \qquad \forall v_h \in H_h \ . \tag{7.39}$$

If u solves (7.37), then Theorem 7.14 implies that

$$\|u - u_h\|_{H^1} \leq C \min_{v_h \in H_h} \|u - v_h\|_{H^1} \leq C \|u\|_{H^2} h \leq C \|f\|_{L_2} h \ ,$$

using elliptic regularity. That is, the finite element approximations converge to the true solution linearly in the grid spacing h.

The problem (7.39) is easily solved, e.g., by computer, since it reduces to a problem in linear algebra. For each $i = 1, 2, ..., n-1$, let $\phi_{h,i} \in H_h$ be such that

$$\phi_{h,i}(x_j) = \begin{cases} 0 & \text{if } i \neq j \ , \\ 1 & \text{if } i = j \ . \end{cases}$$

Then $\{\phi_{h,i}\}_{i=1}^{n-1}$ forms a vector space basis for H_h, and so there are coefficients $\alpha_i \in \mathbb{R}$ such that

$$u_h(x) = \sum_{j=1}^{n-1} \alpha_j \phi_{h,j}(x) \ ,$$

and (7.39) reduces to

$$\sum_{j=1}^{n-1} \alpha_j (\phi'_{h,j}, \phi'_{h,i})_{L_2} = (f, \phi_{h,i})_{L_2} \qquad \forall i = 1, 2, ..., n-1 \ ,$$

since it is sufficient to test against the basis functions $\phi_{h,i}$. Let the $(n-1) \times (n-1)$ matrix M be defined by

$$M_{i,j} = (\phi'_{h,j}, \phi'_{h,i})_{L_2}$$

and the $(n-1)$-vectors a and b by

$$a_j = \alpha_j \qquad \text{and} \qquad b_i = (f, \phi_{h,i})_{L_2} \ .$$

Then our problem is simply $Ma = b$, and the coefficients of u_h are given from the solution $a = M^{-1}b$ (why is this matrix invertible?).

7.7. Green's functions

Let $\mathcal{L}$ be a linear partial differential operator, such as is given in (7.3). Often we can find a fundamental solution $E \in \mathcal{D}'$ satisfying

$$\mathcal{L}E = \delta_0 \ ,$$

wherein δ_0 is the Dirac delta function or point mass at the origin. If for the moment we consider that $\mathcal{L}$ has constant coefficients, then we know from the Malgrange and Ehrenpreis Theorem 4.27, that such a fundamental solution exists. It is not unique, but for $f \in \mathcal{D}$, say, the equation $\mathcal{L}u = f$ has a solution $u = E * f$. However, u, defined this way, will generally fail to satisfy any imposed boundary condition. To resolve this difficulty, we define a special fundamental solution in this section. For maximum generality, we will often proceed formally, assuming sufficient smoothness of all quantities involved to justify the calculations.

Let $\mathcal{B}$ denote a linear boundary condition operator (which generally involves the traces γ_0 and/or γ_1, and represents a Dirichlet, Neumann, or Robin boundary condition). For reasonable f and g, we consider the BVP

$$\begin{cases} \mathcal{L}u = f & \text{in } \Omega \ , \\ \mathcal{B}u = g & \text{on } \partial\Omega \ . \end{cases} \tag{7.40}$$

Initially we will consider the homogeneous case where $g = 0$.

Definition. Suppose $\Omega \subset \mathbb{R}^d$, $\mathcal{L}$ is a linear partial differential operator, and $\mathcal{B}$ is a homogeneous linear boundary condition. We call $G : \Omega \times \Omega \to \mathbb{R}$ a *Green's function* for $\mathcal{L}$ and $\mathcal{B}$ if, for any $f \in \mathcal{D}$, a weak solution u of (7.40) with $g = 0$ is given by

$$u(x) = \int_\Omega G(x,y)\, f(y)\, dy \ . \tag{7.41}$$

We assume here that $\partial\Omega$ is smooth enough to support the definition of the boundary condition.

Proposition 7.15. *The Green's function $G(\cdot, y) : \Omega \to \mathbb{R}$ is a fundamental solution for $\mathcal{L}$ with the point mass $\delta_y(\cdot) = \delta_0(\cdot - y)$: for a.e. $y \in \Omega$,*

$$\mathcal{L}_x G(x,y) = \delta_0(x-y) \qquad \text{for } x \in \Omega$$

(wherein we indicate that $\mathcal{L}$ acts on the variable x by writing $\mathcal{L}_x$ instead). Moreover, $G(x,y)$ satisfies the homogeneous boundary condition

$$\mathcal{B}_x G(x,y) = 0 \qquad \text{for } x \in \partial\Omega \ .$$

Proof. For any $f \in \mathcal{D}$, we have u defined by (7.41), which solves $\mathcal{L}u = f$. We would like to calculate

$$f(x) = \mathcal{L}u(x) = \mathcal{L}\int_\Omega G(x,y)\, f(y)\, dy = \int_\Omega \mathcal{L}_x G(x,y)\, f(y)\, dy \ ,$$

which would indicate the result, but we need to justify moving $\mathcal{L}$ inside the integral. So for $\phi \in \mathcal{D}(\Omega)$,

$$\begin{aligned}
\int_\Omega f(x)\, \phi(x)\, dx &= \int_\Omega \mathcal{L}u(x)\, \phi(x)\, dx \\
&= \int_\Omega u(x)\, \mathcal{L}^*\phi(x)\, dx \\
&= \int_\Omega \int_\Omega G(x,y)\, f(y)\, \mathcal{L}^*\phi(x)\, dy\, dx \\
&= \int_\Omega \int_\Omega G(x,y)\, \mathcal{L}^*\phi(x)\, f(y)\, dx\, dy \\
&= \int_\Omega \langle \mathcal{L}_x G(\cdot, y), \phi\rangle\, f(y)\, dy \ ,
\end{aligned}$$

showing that

$$\langle \mathcal{L}_x G(\cdot, y), \phi\rangle = \phi(y) \ ,$$

that is, $\mathcal{L}_x G(x,y) = \delta_y(x)$.

That $G(x,y)$ satisfies a homogeneous Dirichlet condition in x is clear. Other boundary conditions involve normal derivatives, and it can be shown as above that G must satisfy them. $\square$

Remark. For a fundamental solution of a constant coefficient operator, $\mathcal{L}E = \delta_0$, translation implies that

$$\mathcal{L}_x E(x-y) = \delta_y(x) \ ,$$

which can be understood as giving the response of the operator at $x \in \mathbb{R}^d$, $E(x-y)$, to a point disturbance δ_y at $y \in \mathbb{R}^d$. Multiplying by the weight $f(y)$ and integrating (i.e., adding the responses) gives the solution $u = E * f$. When boundary conditions are imposed, a point disturbance at y is not necessarily translation equivalent to a disturbance at $\tilde{y} \neq y$. This is also true of nonconstant coefficient operators. Thus the more general form of the Green's function being a function of two variables is required: $G(x,y)$ is the response of the operator at $x \in \Omega$ to a point disturbance at $y \in \Omega$, subject also to the boundary conditions.

Given a fundamental solution E that is sufficiently smooth outside the origin, we can construct the Green's function by solving a related BVP. For almost every $y \in \Omega$, solve

$$\begin{cases} \mathcal{L}_x w_y(x) = 0 & \text{for } x \in \Omega \ , \\ \mathcal{B}_x w_y(x) = \mathcal{B}_x E(x-y) & \text{for } x \in \partial\Omega \ , \end{cases}$$

and then

$$G(x,y) = E(x-y) - w_y(x)$$

is the Green's function. Note that indeed $\mathcal{L}_x G(x,y) = \delta_0(x-y)$ is a fundamental solution, and that this one is special in that $\mathcal{B}_x G(x,y) = 0$ on $\partial\Omega$.

It is generally difficult to find an explicit expression for the Green's function, except in special cases. However, its existence implies that the inverse operator of $(\mathcal{L}, \mathcal{B})$ is an integral operator, and thus has many important properties, such as compactness. When G can be found explicitly, it can be a powerful tool both theoretically and computationally.

We now consider the nonhomogeneous BVP (7.40). Suppose that there is u_0 defined in Ω such that $\mathcal{B}u_0 = g$ on $\partial\Omega$. Then, if $w = u - u_0$,

$$\begin{cases} \mathcal{L}w = f - \mathcal{L}u_0 & \text{in } \Omega \ , \\ \mathcal{B}w = 0 & \text{on } \partial\Omega \ , \end{cases} \tag{7.42}$$

and this problem has a Green's function $G(x,y)$. Thus our solution is

$$u(x) = w(x) + u_0(x) = \int_\Omega G(x,y)\big(f(y) - \mathcal{L}u_0(y)\big)\, dy + u_0(x) \ .$$

This formula has limited utility, since we cannot easily find u_0.

In some cases, the Green's function can be used to define a different integral operator involving an integral on $\partial\Omega$ which involves g directly. To illustrate, consider (7.40) with $\mathcal{L} = -\Delta + I$, where I is the identity operator. Now $\mathcal{L}_x G(x,y) = \delta_y(x)$, so this fact and integration by parts implies that

$$\begin{aligned} u(y) &= \int_\Omega \mathcal{L}_x G(x,y)\, u(x)\, dx \\ &= \int_\Omega G(x,y)\, u(x)\, dx + \int_\Omega \nabla_x G(x,y) \cdot \nabla u(x)\, dx - \int_{\partial\Omega} \nabla_x G(x,y) \cdot \nu\, u(x)\, d\sigma(x) \\ &= \int_\Omega G(x,y)\, \mathcal{L}u(x)\, dx + \int_{\partial\Omega} G(x,y) \nabla u(x) \cdot \nu\, d\sigma(x) - \int_{\partial\Omega} \nabla_x G(x,y) \cdot \nu\, u(x)\, d\sigma(x) \ . \end{aligned}$$

If $\mathcal{B}$ imposes the Dirichlet BC, so $u = u_D$, then since $\mathcal{L}u = f$ and $G(x,y)$ itself satisfies the homogeneous boundary conditions in x, we have simply

$$u(y) = \int_\Omega G(x,y)\, f(x)\, dx - \int_{\partial\Omega} \nabla_x G(x,y) \cdot \nu\, u_D(x)\, d\sigma(x) \ .$$

This is called the *Poisson integral formula.* If instead $\mathcal{B}$ imposes the Neumann BC, so $-\nabla u \cdot \nu = g$, then

$$u(x) = \int_\Omega G(x,y)\, f(x)\, dx - \int_{\partial\Omega} G(x,y)\, g(x)\, d\sigma(x) \ .$$

We remark that when a compatibility condition condition is required, it is not always possible to obtain the Green's function directly. For example, if $\mathcal{L} = -\Delta$ and we have the nonhomogeneous Neumann problem, then $\int_\Omega \delta_y(x)\,dx = 1 \neq 0$ as is required. So, instead we solve

$$\begin{cases} \quad -\Delta_x G(x,y) = \delta_y(x) - 1/|\Omega| & \text{in } \Omega\ , \\ -\nabla_x G(x,y)\cdot\nu = 0 & \text{on } \partial\Omega\ , \end{cases}$$

where $|\Omega|$ is the measure of Ω. Then our BVP (7.40) has the extra condition that the average of u vanishes. Thus, as above,

$$\begin{aligned} u(y) &= -\int_\Omega \Delta_x G(x,y)\,u(x)\,dx \\ &= \int_\Omega \nabla_x G(x,y)\cdot\nabla u(x)\,dx - \int_{\partial\Omega} \nabla_x G(x,y)\cdot\nu\,u(x)\,d\sigma(x) \\ &= -\int_\Omega G(x,y)\,\Delta u(x)\,dx + \int_{\partial\Omega} G(x,y)\nabla u(x)\cdot\nu\,d\sigma(x) \\ &= \int_\Omega G(x,y)\,f(x)\,dx - \int_{\partial\Omega} G(x,y)\,g(x)\,d\sigma(x)\ . \end{aligned}$$

7.8. Exercises

1. If A is a positive definite matrix, show that its eigenvalues are positive. Conversely, prove that if A is symmetric and has positive eigenvalues, then A is positive definite.

2. Suppose that the hypotheses of the Generalized Lax-Milgram Theorem 7.10 are satisfied. Suppose also that $x_{0,1}$ and $x_{0,2}$ are in $\mathcal{X}$ are such that the sets $X + x_{0,1} = X + x_{0,2}$. Prove that the solutions $u_1 \in X + x_{0,1}$ and $u_2 \in X + x_{0,2}$ of the abstract variational problem (7.18) agree (i.e., $u_1 = u_2$). What does this result say about Dirichlet boundary value problems?

3. Suppose that we wish to find $u \in H^2(\Omega)$ solving the nonlinear problem $-\Delta u + cu^2 = f \in L_2(\Omega)$, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain. For consistency, we would require that $cu^2 \in L_2(\Omega)$. Determine the smallest p such that if $c \in L_p(\Omega)$, you can be certain that this is true, if indeed it is possible. The answer depends on d.

4. Suppose $\Omega \subset \mathbb{R}^d$ is a connected Lipschitz domain and $V \subset \Omega$ has positive measure. Let $H = \{u \in H^1(\Omega) : u|_V = 0\}$.

   (a) Why is H a Hilbert space?

   (b) Prove the following Poincaré inequality: there is some $C > 0$ such that
   $$\|u\|_{L_2(\Omega)} \le C\|\nabla u\|_{L_2(\Omega)} \qquad \forall u \in H\ .$$

5. Suppose that $\Omega \subset \mathbb{R}^d$ is a smooth, bounded, connected domain. Let
   $$H = \left\{ u \in H^2(\Omega) : \int_\Omega u(x)\,dx = 0 \text{ and } \nabla u\cdot\nu = 0 \text{ on } \partial\Omega \right\}\ .$$
   Show that H is a Hilbert space, and prove that there exists $C > 0$ such that for any $u \in H$,
   $$\|u\|_{H^1(\Omega)} \le C \sum_{|\alpha|=2} \|D^\alpha u\|_{L_2(\Omega)}\ .$$

6. Suppose $\Omega \subset \mathbb{R}^d$ is a $C^{1,1}$ domain. Consider the biharmonic BVP

$$\begin{cases} \Delta^2 u = f & \text{in } \Omega\ , \\ \nabla u \cdot \nu = g & \text{on } \partial\Omega\ , \\ u = u_D & \text{on } \partial\Omega\ , \end{cases}$$

wherein $\Delta^2 u = \Delta\Delta u$ is the application of the Laplace operator twice.

(a) Determine appropriate Sobolev spaces within which the functions u, f, g, and u_D should lie, and formulate an appropriate variational problem for the BVP. Show that the two problems are equivalent.

(b) Show that there is a unique solution to the variational problem. [Hint: use the Elliptic Regularity Theorem to prove coercivity of the bilinear form.]

(c) What are the natural BC's for this problem?

(d) For simplicity, let u_D and g vanish and define the energy functional

$$J(v) = \int_\Omega \big(|\Delta v(x)|^2 - 2f(x)\, v(x)\big)\, dx\ ,$$

Prove that minimization of J is equivalent to the variational problem.

7. Suppose $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain. Consider the Stokes problem for vector u and scalar p given by

$$\begin{cases} -\Delta u + \nabla p = f & \text{in } \Omega\ , \\ \nabla \cdot u = 0 & \text{in } \Omega\ , \\ u = 0 & \text{on } \partial\Omega\ , \end{cases}$$

where the first equation holds for each coordinate (i.e., $-\Delta u_j + \partial p/\partial x_j = f_j$ for each $j = 1, ..., d$). This problem is not a minimization problem; rather, it is a *saddle-point problem*, in that we minimize some energy subject to the *constraint* $\nabla \cdot u = 0$. However, if we work over the constrained space, we can handle this problem by the ideas of this chapter. Let

$$H = \{v \in (H_0^1(\Omega))^d : \nabla \cdot u = 0\}\ .$$

(a) Verify that H is a Hilbert space.

(b) Determine an appropriate Sobolev space for f, and formulate an appropriate variational problem for the constrained Stokes problem.

(c) Show that there is a unique solution to the variational problem.

8. Show that for $f \in L_2(\mathbb{R}^d)$, there exists a unique solution $u \in H^1(\mathbb{R}^d)$ of the boundary value problem

$$\begin{cases} -\Delta u + u = f & \text{in } \mathbb{R}^d\ , \\ u \to 0 & \text{as } |x| \to \infty\ . \end{cases}$$

9. Consider the boundary value problem for $u(x, y) : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\begin{aligned} -\,u_{xx} + e^y u = f\ , \qquad & \text{for } (x,y) \in (0,1)^2\ , \\ u(0,y) = 0,\ u(1,y) = \cos(y)\ , \qquad & \text{for } y \in (0,1)\ . \end{aligned}$$

Rewrite this as a variational problem and show that there exists a unique solution. Be sure to define your function spaces carefully and identify where f must lie.

10. Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with a Lipschitz boundary, $f \in L_2(\Omega)$, and $\alpha > 0$. Consider the Robin boundary value problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \ , \\ \dfrac{\partial u}{\partial \nu} + \alpha\, u = 0 & \text{on } \partial\Omega \ . \end{cases}$$

(a) For this problem, formulate a variational principle

$$B(u, v) = (f, v) \qquad \forall v \in H^1(\Omega) \ .$$

(b) Show that this problem has a unique weak solution.

11. Let $\Omega = [0, 1]^d$, define

$$H^1_{\#}(\Omega) = \left\{ v \in H^1(\Re^d) \ : \ v \text{ is periodic of period 1 in each direction and } \int_\Omega v\, dx = 0 \right\} ,$$

and consider the problem of finding a periodic solution $u \in H^1_{\#}(\Omega)$ of

$$-\Delta u = f \quad \text{on } \Omega \ ,$$

where $f \in L_2(\Omega)$.

(a) Define precisely what it means for $v \in H^1(\Re^d)$ to be periodic of period 1 in each direction.

(b) Show that $H^1_{\#}(\Omega)$ is a Hilbert space.

(c) Show that there is a unique solution to the partial differential equation.

12. Consider

$$B(u, v) = (a\nabla u, \nabla v)_{L_2(\Omega)} + (bu, \nabla v)_{L_2(\Omega)} + (cu, v)_{L_2(\Omega)}$$

(a) Derive a condition on b to insure that B is coercive on $H^1(\Omega)$ when a is uniformly positive definite *and* c is uniformly positive.

(b) Suppose $b = 0$. If $c < 0$, is B not coercive? Show that this is true on $H^1(\Omega)$, but that by restricting how negative c may be, B is still coercive on $H^1_0(\Omega)$.

13. Modify the statement of Theorem 7.14 to allow for nonhomogeneous essential boundary conditions, and prove the result.

14. Consider the finite element method in Section 7.6.

(a) Modify the method to account for nonhomogeneous Neumann conditions.

(b) Modify the method to account for nonhomogeneous Dirichlet conditions.

15. Compute explicitly the finite element solution to (7.36) using $f(x) = x^2(1 - x)$ and $n = 4$. How does this approximation compare to the true solution?

16. Let H_h be the set of continuous piecewise linear functions defined on the grid $x_j = jh$, where $h = 1/n$ for some integer $n > 0$. Let the interpolation operator $\mathcal{I}_h : H^1_0(0, 1) \to H_h$ be defined by

$$\mathcal{I}_h v(x_j) = v(x_j) \qquad \forall j = 1, 2, ..., n - 1 \ .$$

(a) Show that $\mathcal{I}_h$ is well defined, and that it is continuous. [Hint: use the Sobolev Imbedding Theorem.]

(b) Show that there is a constant $C > 0$ independent of h such that

$$\|v - \mathcal{I}_h v\|_{H^1(x_{j-1},x_j)} \leq C\|v\|_{H^2(x_{j-1},x_j)}h \ .$$

[Hint: change variables so that the domain becomes $(0, 1)$, where the result is trivial by Poincaré's inequality.]

(c) Show that (7.38) holds.

17. Consider the problem (7.36).

(a) Find the Green's function.

(b) Instead impose Neumann BC's, and find the Green's function. [Hint: recall that now we require $-(\partial^2/\partial x^2)G(x, y) = \delta_y(x) - 1$.]

CHAPTER 8

Differential Calculus in Banach Spaces and the Calculus of Variations

In this chapter, we move away from the rigid, albeit very useful confines of linear maps and consider maps $f : U \to Y$, not necessarily linear, where U is an open set in a Banach space X and Y is also a Banach space.

As in finite-dimensional calculus, we begin the analysis of such functions by effecting a local approximation. In one-variable calculus, we are used to writing

$$f(x) \cong f(x_0) + f'(x_0)(x - x_0) \tag{8.1}$$

when $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, say. This amounts to approximating f by an affine function, a translation of a linear mapping. This procedure allows the method of linear functional analysis to be brought to bear upon understanding a nonlinear function f.

8.1. Differentiation

In attempting to generalize the notion of a derivative to more than one dimension, one realizes immediately that the one-variable calculus formula

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{8.2}$$

cannot be taken over intact. First, the quantity $1/h$ has no meaning in higher dimensions. Secondly, whatever $f'(x)$ might be, it is plainly not going to be a number. Instead, just as in multivariable calculus, it is a precise version of (8.1) that readily generalizes, and not (8.2). We digress briefly for a definition.

Definition. Suppose X, Y are NLS's and $f : X \to Y$. If

$$\frac{\|f(h)\|_Y}{\|h\|_X} \to 0 \quad \text{as } h \to 0 \ ,$$

we say that as h tends to 0, f is "little oh" of h, and we denote this as

$$\|f(h)\|_Y = o(\|h\|_X) \ .$$

Definition. Let $f : U \to Y$ where $U \subset X$ is open and X and Y are normed linear spaces. Let $x \in U$. We say that f is *Fréchet-differentiable* at x if there is an element $A \in B(X, Y)$ such that if

$$R(x, h) = f(x + h) - f(x) - Ah \ , \tag{8.3}$$

then

$$\frac{1}{\|h\|_X} \|R(x, h)\|_Y \to 0 \tag{8.4}$$

as $h \to 0$ in X, i.e.,

$$\|R(x,h)\|_Y = o(\|h\|_X) \ .$$

When it exists, we call A the Fréchet-derivative of f at x; it is denoted variously by

$$A = A_x = f'(x) = Df(x) \ . \tag{8.5}$$

Notice that this generalizes the one-dimensional idea of being differentiable. Indeed, if $f \in C^1(\mathbb{R})$, then

$$R(x,h) = f(x+h) - f(x) - f'(x)h = \left[\frac{f(x+h)-f(x)}{h} - f'(x)\right] h \ ,$$

and so

$$\frac{|R(x,h)|}{|h|} = \left|\frac{f(x+h)-f(x)}{h} - f'(x)\right| \to 0$$

as $h \to 0$ in $\mathbb{R}$. Note that $B(\mathbb{R},\mathbb{R}) = \mathbb{R}$, and thus that the product $f'(x)h$ may be viewed as the linear mapping that sends h to $f'(x)h$.

We can also think of Df as a mapping of $X \times X$ into Y via the correspondence

$$(x,h) \longmapsto f'(x)h \ .$$

Proposition 8.1. *If f is Fréchet differentiable, then $Df(x)$ is unique and f is continuous at x.*

Proof. Suppose $A, B \in B(X,Y)$ are such that

$$f(x+h) - f(x) - Ah = R_A(x,h)$$

and

$$f(x+h) - f(x) - Bh = R_B(x,h) \ ,$$

where

$$\frac{\|R_A(x,h)\|_Y}{\|h\|_X} \to 0 \quad \text{and} \quad \frac{\|R_B(x,h)\|_Y}{\|h\|_X} \to 0$$

as $h \to 0$ in X. It follows that

$$\begin{aligned}
\|A-B\|_{B(X,Y)} &= \frac{1}{\varepsilon} \sup_{\|h\|_X=\varepsilon} \|Ah - Bh\|_Y \\
&= \sup_{\|h\|_X=\varepsilon} \frac{\|R_B(x,h) - R_A(x,h)\|_Y}{\|h\|_X} \\
&\leq \sup_{\|h\|_X=\varepsilon} \frac{\|R_B(x,h)\|_Y}{\|h\|_X} + \sup_{\|h\|_X=\varepsilon} \frac{\|R_A(x,h)\|_Y}{\|h\|_X} \ ,
\end{aligned}$$

and the right-hand side may be made as small as we like by taking ε small enough. Thus $A = B$.

Continuity of f at x is straightforward since

$$\begin{aligned}
\|f(x+h) - f(x)\|_Y &= \|Df(x)h + R(x,h)\|_Y \\
&\leq \|Df(x)\|_{B(X,Y)} \|h\|_Y + \|R(x,h)\|_Y \ ,
\end{aligned}$$

and the right-hand side tends to 0 as $h \to 0$ in X. □

In fact, we have much more than mere continuity. The following result is often useful. It says that when f is differentiable, it is locally Lipschitz.

LEMMA 8.2 (Local-Lipschitz property). *If $f : U \to Y$ is differentiable at $x \in U$, then given $\varepsilon > 0$, there is a $\delta = \delta(x, \varepsilon) > 0$ such that for all h with $\|h\|_X \le \delta$,*

$$\|f(x+h) - f(x)\|_Y \le \Big(\|Df(x)\|_{B(X,Y)} + \varepsilon\Big)\|h\|_X \ . \tag{8.6}$$

PROOF. Simply write

$$f(x+h) - f(x) = R(x,h) + Df(x)h \ . \tag{8.7}$$

Since f is differentiable at x, given $\varepsilon > 0$, there is a $\delta > 0$ such that $\|h\|_X \le \delta$ implies

$$\frac{\|R(x,h)\|_Y}{\|h\|_X} \le \varepsilon \ .$$

Then (8.7) implies the advertised results. □

EXAMPLES. 1. If $f(x) = Bx$, where $B \in B(X,Y)$, then f is Fréchet-differentiable everywhere and

$$Df(x) = B$$

for all $x \in X$.

2. Let $X = H$ be a Hilbert-space over $\mathbb{R}$. Let $f(x) = (x, Ax)_H$ where $A \in B(H,H)$. Then, $f : H \to \mathbb{R}$ and

$$\begin{aligned} f(x+h) - f(x) &= (x, Ah)_H + (h, Ax)_H + (h, Ah)_H \\ &= \Big((A^* + A)x, h\Big)_H + (h, Ah)_H \ . \end{aligned}$$

Hence if we define, for x, $h \in X$,

$$Df(x)h = \Big((A^* + A)x, h\Big)_H \ ,$$

then

$$\|f(x+h) - f(x) - Df(x)h\|_Y \le \|h\|_X^2 \|A\|_{B(X,Y)} \ .$$

Thus $Df(x) \in H^* = B(H, \mathbb{R})$ is the Riesz-map associated with the element $(A^* + A)x$.

3. Let $f : \mathbb{R}^n \to \mathbb{R}$ and suppose $f \in C^1(\mathbb{R}^n)$, which is to say $\partial_i f$ exists and is continuous on $\mathbb{R}^n$, $1 \le i \le n$. Then $Df(x) \in B(\mathbb{R}^n, \mathbb{R})$ is defined by

$$Df(x)h = \nabla f(x) \cdot h \ .$$

4. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ and suppose $f \in C^1(\mathbb{R}^n, \mathbb{R}^m)$, which is to say each of the component functions $f = (f_1, \dots, f_m)$ as a $\mathbb{R}$-valued function, having all its first partial derivatives, and each of these is continuous. Then f is Fréchet-differentiable and

$$Df(x)h = [\partial_j f_i(x)]h \ ,$$

where the latter is matrix multiplication and the matrix itself is the usual Jacobian matrix. That is, $Df(x) \in B(\mathbb{R}^n, \mathbb{R}^m)$ is an $m \times n$ matrix, and the ith component of $Df(x)h$ is

$$\sum_{j=1}^{n} \partial_j f_i(x) h_j \ .$$

5. Let $\varphi \in L_p(\mathbb{R}^n)$, where $p \geq 1$, p an integer, and define

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

Then $f : L_p(\mathbb{R}^n) \to \mathbb{R}$, f is Fréchet-differentiable and

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

There is a differentiability notion weaker than Fréchet-differentiable, but still occasionally useful. In this conception, we only ask the function f to be differentiable in a specified direction. Let $h \in X$ and consider the Y-valued function of the real variable t:

$$g(t) = f(x + th) \ .$$

Definition. Suppose $f : X \to Y$. Then f is *Gateaux-differentiable* at x in the *direction* $h \in X$ if there is an $A \in B(X, Y)$ such that

$$\frac{1}{t}\|f(x + th) - f(x) - tAh\| \to 0$$

as $t \to 0$. The Gateaux-derivative is denoted by

$$A = D_h f(x) \ .$$

Moreover, f is *Gateaux-differentiable* at x if it is Gateaux-differentiable at x in every direction $h \in X$.

Proposition 8.3. *If f is Fréchet-differentiable, then it is Gateaux-differentiable.*

Remark. The converse is not valid. The function $f : \mathbb{R}^2 \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 0 \ , & \text{if } x_2 = 0 \ , \\ x_1^3/x_2 \ , & \text{if } x_2 \neq 0 \ , \end{cases}$$

is not continuous at the origin. For instance $f((t, t^3)) \to 1$ as $t \to 0$, but $f(0) = 0$. However, f is Gateaux-differentiable at $(0, 0)$ in every direction h since

$$\frac{f(th) - f(0)}{t} = \frac{f(th)}{t} = \begin{cases} 0 & \text{if } h_2 = 0 \ , \\ t(h_1^3/h_2) & \text{if } h_2 \neq 0 \ . \end{cases}$$

The limit as $t \to 0$ exists and is zero, whatever the value of h.

Theorem 8.4 (Chain Rule). *Let X, Y, Z be NLS's and $U \subset X$ open, $V \subset Y$ open, $f : U \to Y$ and $g : V \to Z$. Let $x \in U$ and $y = f(x) \in V$. Suppose g is Fréchet-differentiable at y and f is Gateaux- (respectively, Fréchet-) differentiable at x. Then $g \circ f$ is Gateaux- (respectively, Fréchet-) differentiable at x and*

$$D(g \circ f)(x) = Dg(y) \circ Df(x) \ .$$

Proof. The proof is given for the case where both maps are Fréchet differentiable. The proof for the Gateaux case is similar. Write

$$R_f(x, h) = f(x + h) - f(x) - Df(x)h$$

and

$$R_g(y, k) = g(y + k) - g(y) - Dg(y)k \ .$$

By assumption,

$$\frac{R_f(x,h)}{\|h\|} \xrightarrow{Y} 0 \quad \text{as} \quad h \xrightarrow{X} 0 \tag{8.8}$$

and

$$\frac{R_g(y,k)}{\|k\|} \xrightarrow{Z} 0 \quad \text{as} \quad k \xrightarrow{Y} 0 \ . \tag{8.9}$$

Define

$$u = u(h) = f(x+h) - f(x) = f(x+h) - y \ . \tag{8.10}$$

By continuity, $u(h) \to 0$ as $h \to 0$. Now consider the difference

$$\begin{aligned} g(f(x+h)) - g(f(x)) &= g(f(x+h)) - g(y) \\ &= Dg(y)[f(x+h) - y] + R_g(y,u) \\ &= Dg(y)[Df(x)h + R_f(x,h)] + R_g(y,u) \\ &= Dg(y)Df(x)h + R(x,h) \ , \end{aligned}$$

where

$$R(x,h) = Dg(y)R_f(x,h) + R_g(y,u) \ .$$

We must show that $R(x,h) = o(\|h\|_X)$ as $h \to 0$. Notice that

$$\frac{\|Dg(y)R_f(x,h)\|_Z}{\|h\|_X} \le \|Dg(y)\|_{B(Y,Z)} \frac{\|R_f(x,h)\|}{\|h\|_X} \to 0 \quad \text{as} \quad h \to 0$$

because of (8.8). The second term is slightly more interesting. We are trying to show

$$\frac{\|R_g(y,u)\|_Z}{\|h\|_X} \to 0 \tag{8.11}$$

as $h \to 0$. This does not follow immediately from (8.9). However, the local-Lipschitz property comes to our rescue.

If $u = 0$, then $R_g(y,u) = 0$. If not, then multiply and divide by $\|u\|_Y$ to reach

$$\frac{\|R_g(y,u)\|_Z}{\|h\|_X} = \frac{\|R_g(y,u)\|}{\|u\|_Y} \ \frac{\|u\|_Y}{\|h\|_X} \ . \tag{8.12}$$

Let $\varepsilon > 0$ be given and suppose without loss of generality that $\varepsilon \le 1$. There is a $\sigma > 0$ such that if $\|k\|_Y \le \sigma$, then

$$\frac{\|R_g(y,k)\|_Z}{\|k\|_Y} \le \varepsilon \ . \tag{8.13}$$

On the other hand, because of (8.6), there is a $\delta > 0$ such that $\|h\|_X \le \delta$ implies

$$\|u(h)\|_Y = \|f(x+h) - f(x)\|_Y \le \Big(\|Df(x)\|_{B(X,Y)} + 1\Big)\|h\|_X \le \sigma \tag{8.14}$$

(simply choose δ so that $\delta(\|Df(x)\|_{B(X,Y)} + 1) \le \sigma$ in addition to it satisfying the smallness requirement in Lemma 8.2). With this choice of δ, if $\|h\|_X \le \delta$, then (8.12) implies

$$\frac{\|R_g(y,u)\|_Z}{\|h\|_X} \le \varepsilon\Big(\|Df(x)\|_{B(X,Y)} + 1\Big) \ .$$

The result follows. $\square$

PROPOSITION 8.5 (Mean-Value Theorem for Curves). *Let Y be a NLS and $\varphi : [a,b] \to Y$ be continuous, where $a < b$ are real numbers. Suppose $\varphi'(t)$ exists on (a,b) and that $\|\varphi'(t)\|_{B(\mathbb{R},Y)} \le M$. Then*

$$\|\varphi(b) - \varphi(a)\|_Y \le M(b-a) \ . \tag{8.15}$$

REMARK. Every bounded linear operator from $\mathbb{R}$ to Y is given by $t \mapsto ty$ for some fixed $y \in Y$. Hence we may identify $\varphi'(t)$ with this element y. Notice in this case that y can be obtained by the elementary limit

$$y = \varphi'(t) = \lim_{s\to 0} \frac{\varphi(t+s) - \varphi(t)}{s} \ .$$

PROOF. Fix an $\varepsilon > 0$ and suppose $\varepsilon \le 1$. For any $t \in (a,b)$, there is a $\delta_t = \delta(t,\varepsilon)$ such that if $|s-t| < \delta_t$, then

$$\|\varphi(s) - \varphi(t)\|_Y < (M+\varepsilon)|s-t| \tag{8.16}$$

by the Local-Lipschitz Lemma 8.2). Let

$$\tilde{S}(t) = \{s \in [a,b] : (8.16) \text{ holds}\} \cup \{t\} \ ,$$

which is open by continuity of φ. Let $S(t)$ be the connected component of $\tilde{S}(t)$ containing t. Then if $a < \tilde{a} < \tilde{b} < b$,

$$[\tilde{a},\tilde{b}] \subset \bigcup_{t\in[\tilde{a},\tilde{b}]} S(t) \ .$$

The sets $S(t)$ are open, being connected components of the open set $\tilde{S}(t)$. Hence by compactness, there is a finite sub-cover, say $S(\tilde{a}), S(\tilde{t}_1), \dots, S(\tilde{b})$ where $\tilde{a} < \tilde{t}_1 < \cdots < \tilde{b}$. This allows us to form a partition of $[\tilde{a},\tilde{b}]$, into N intervals, say

$$\tilde{a} = t_0 < t_2 < \cdots < t_{2N} = \tilde{b} \ ,$$

in such a way that $S(t_{2k+2}) \cap S(t_{2k}) \ne \emptyset$ for all k. Choose points $t_{2k+1} \in S(t_{2k+2}) \cap S(t_{2k})$, enrich the partition to

$$\tilde{a} = t_0 < t_1 < t_2 < \cdots < t_{2N} = \tilde{b} \ ,$$

and note that

$$\|\varphi(t_{k+1}) - \varphi(t_k)\|_Y \le (M+\varepsilon)|t_{k+1} - t_k|$$

for all k. Hence

$$\begin{aligned}\|\varphi(\tilde{b}) - \varphi(\tilde{a})\|_Y &\le \sum_{k=1}^{2N} \|\varphi(t_k) - \varphi(t_{k-1})\|_Y \\ &\le (M+\varepsilon)\sum_{k=1}^{2N}(t_k - t_{k-1}) = (M+\varepsilon)(\tilde{b} - \tilde{a}) \ .\end{aligned}$$

By continuity, we may take the limit on $\tilde{b} \to b$ and $\tilde{a} \to a$, and the same inequality holds. Since $\varepsilon > 0$ was arbitrary, (8.15) follows. $\square$

REMARK. The Mean-Value Theorem for curves can be used to give reasonable conditions under which Gateaux-differentiability implies Fréchet-differentiability. Here is another corollary of this result.

THEOREM 8.6 (Mean-Value Theorem). *Let X, Y be NLS's and $U \subset X$ open. Let $f : U \to Y$ be Fréchet-differentiable everywhere in U and suppose the line segment*

$$\ell = \{tx_2 + (1-t)x_1 : 0 \le t \le 1\}$$

is contained in U. Then

$$\|f(x_2) - f(x_1)\|_Y \le \sup_{x \in \ell} \|Df(x)\|_{B(X,Y)} \|x_2 - x_1\|_X \ . \tag{8.17}$$

PROOF. Define $\varphi : [0,1] \to Y$ by

$$\varphi(t) = f((1-t)x_1 + tx_2) = f(x_1 + t(x_2 - x_1)) = f(\gamma(t)) \ ,$$

where $\gamma : [0,1] \to X$. Certainly φ is differentiable on $[0,1]$ by the chain rule. By Proposition 8.5,

$$\|f(x_2) - f(x_1)\|_Y = \|\varphi(1) - \varphi(0)\|_Y \le \sup_{0 \le t \le 1} \|\varphi'(t)\|_Y \ .$$

but, the chain rule insures that

$$\varphi'(t) = Df(\gamma(t)) \circ \gamma'(t) = Df(\gamma(t))(x_2 - x_1) \ ,$$

so

$$\begin{aligned} \|\varphi'(t)\|_Y &\le \|Df(\gamma(t))\|_{B(X,Y)} \|x_2 - x_1\|_X \\ &\le \sup_{x \in \ell} \|Df(x)\|_{B(X,Y)} \|x_2 - x_1\|_X \ . \qquad \square \end{aligned}$$

One can generalize the discussion immediately to partial Fréchet-differentiability. Suppose $X_1, \dots, X_m$ are NLS's over $\mathbb{F}$ and Y another NLS. Let

$$X = X_1 \times \cdots \times X_m$$

be the Cartesian product of the X_i's, let the vector-space operations be carried out componentwise, and let the norm be any of the equivalent functions

$$\|x\|_X = \left(\sum_{j=1}^m \|x_i\|_{X_i}^p \right)^{1/p} = \big\| (\|x_1\|_{X_1}, \cdots, \|x_m\|_{X_m}) \big\|_{\ell_p} \ , \tag{8.18}$$

where $p \in [1, \infty]$ and $x = (x_1, \dots, x_m)$. Of course, X is a Banach space if and only if X_i is a Banach space, $1 \le i \le m$. Conversely, we could begin with a direct sum decomposition

$$X = X_1 \oplus \cdots \oplus X_m \ ,$$

with norms $\| \cdot \|_{X_i} = \| \cdot \|_X$, and associate this with the equivalent Banach space $X_1 \times \cdots \times X_m$.

DEFINITION. Let $X = X_1 \times \cdots \times X_m$ as above. Let $U \subset X$ be open and $F : U \to Y$. Let $x = (x_1, \dots, x_m) \in U$ and fix an integer $k \in [1, m]$. For z near x_k in X_k, the point $(x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_m)$ lies in U, since U is open. Define

$$f_k(z) = F(x_1, \dots, x_{k-1}, z, x_{k+1}, \dots, x_m) \ .$$

Then f_k maps an open subset of X_k into Y. If f_k has a Fréchet derivative at $z = x_k$, then we say F has a kth-partial derivative at x and define

$$D_k F(x) = Df_k(x_k) \ .$$

Notice that $D_k F(x) \in B(X_k, Y)$.

PROPOSITION 8.7. *Let $X = X_1 \times \cdots \times X_m$ be the Cartesian product of NLS's, $U \subset X$ open, and $F : U \to Y$, another NLS. Suppose $D_j F(x)$ exists for $x \in U$ and $1 \le j \le m$, and that these linear maps are continuous as a function of x at $x_0 \in U$. Then F is Fréchet-differentiable at x_0 and for $h = (h_1, \ldots , h_m) \in X$,*

$$DF(x_0)h = \sum_{j=1}^{m} D_j F(x_0)h_j \ . \tag{8.19}$$

PROOF. The right-hand side of (8.19) defines a bounded linear map on X. Indeed, it may be written as

$$Ah = \sum_{j=1}^{m} D_j F(x_0) \circ \Pi_j h$$

where $\Pi_j : X \to X_j$ is the projection on the jth-component. So A is a sum of compositions of bounded operators and so is itself a bounded operator. Define

$$\sigma(h) = F(x_0 + h) - F(x_0) - Ah \ .$$

It suffices to show that $\sigma : X \to Y$ is such that

$$\frac{\sigma(h)}{\|h\|_X} \to 0$$

as $h \to 0$. Let $\varepsilon > 0$ be given. Because F is partially Fréchet-differentiable and A is linear, it follows immediately from the chain rule that σ is partially Fréchet-differentiable in h and

$$D_j \sigma(h) = D_j F(x_0 + h) - D_j F(x_0) \ .$$

Since the partial Fréchet-derivatives are continuous as a function of x at x_0 it follows there is a $\delta > 0$ such that if $\|h^0\|_X \le \delta$, then

$$\|D_j \sigma(h^0)\|_{B(X_j,Y)} \le \varepsilon \quad \text{for} \quad 1 \le j \le m \ . \tag{8.20}$$

On the other hand,

$$\begin{aligned} \|\sigma(h^0)\|_Y \le \|\sigma(h^0) - \sigma(0, h_2^0, \ldots , h_m^0)\|_Y + \|\sigma(0, h_2^0, \ldots , h_m^0) - \sigma(0, 0, h_3^0, \ldots , h_m^0)\|_Y \\ + \cdots + \|\sigma(0, \ldots , 0, h_m^0) - \sigma(0, \ldots , 0)\|_Y \ . \end{aligned} \tag{8.21}$$

Thus, if $\|h\|_X \le \delta$, then by the Mean-Value Theorem applied to the mappings

$$\sigma_j(h_j) = \sigma(0, \ldots , 0, h_j, h_{j+1}^0, \ldots , h_m^0) \ ,$$

it is determined on the basis of (8.20) that

$$\begin{aligned} \|\sigma_j(h_j) - \sigma_j(0)\|_Y &\le \sup_{t\in[0,1]} \|D\sigma_j(th_j)\|_{B(X_j,Y)} \|h_j\|_{X_j} \\ &= \sup_{t\in[0,1]} \|D_j\sigma(0, \ldots , 0, th_j, h_{j+1}^0, \ldots , h_m^0)\|_{B(X_j,Y)} \|h_j\|_{X_j} \\ &\le \varepsilon \|h_j\|_{X_j} \ , \quad \text{for} \quad 1 \le j \le m \ . \end{aligned}$$

Choosing in (8.18) the ℓ_1-norm on X, it follows from (8.21) and the last inequalities that for $\|h^0\|_X < \delta$,

$$\|\sigma(h^0)\|_Y \le \varepsilon \sum_{j=1}^{m} \|h_j\|_{X_j} = \varepsilon \|h\|_X \ .$$

(If another ℓ_p-norm is used in (8.18), we merely get a fixed constant multiple of the right-hand side above.) The result follows. $\square$

8.2. Nonlinear Equations

Developed here are some helpful techniques for understanding when a nonlinear equation has a solution.

DEFINITION. Let (X,d) be a metric space and $T : X \to X$. The mapping T is a *contraction* if there is a θ with $0 \le \theta < 1$ such that

$$d(Tx,Ty) \le \theta d(x,y) \quad \text{for all } x,y \in X \ .$$

A *fixed point* of the mapping T is an $x \in X$ such that $x = Tx$.

A contraction map is a Lipschitz map with Lipschitz constant less than 1. Such maps are also continuous.

THEOREM 8.8 (Banach Contraction-Mapping Principle). *Let (X,d) be a complete metric space and T a contraction mapping of X. Then there is a unique fixed point of T in X.*

PROOF. If there were two fixed points x and y, then

$$d(x,y) = d(Tx,Ty) \le \theta d(x,y) \ ,$$

and since $d(x,y) \ge 0$ and $0 \le \theta < 1$, it follows that $d(x,y) = 0$, whence $x = y$.

For existence of a fixed point, argue as follows. Fix an $x_0 \in X$ and let $x_1 = Tx_0$, $x_2 = Tx_1$ and so on. We claim the sequence $\{x_n\}_{n=0}^\infty$ of iterates is a Cauchy sequence.

If this $\{x_n\}_{n=0}^\infty$ is Cauchy, then since (X,d) is complete, there is an $\bar{x}$ such that $x_n \to \bar{x}$. But then $Tx_n \to T\bar{x}$ by continuity. Since $Tx_n = x_{n+1}$, it follows that $T\bar{x} = \bar{x}$.

To see $\{x_n\}_{n=0}^\infty$ is a Cauchy sequence, first notice that

$$d(x_1,x_2) = d(Tx_0,Tx_1) \le \theta d(x_0,x_1) \ .$$

Continuing in this manner,

$$d(x_n,x_{n+1}) = d(Tx_{n-1},Tx_n) \le \theta d(x_{n-1},x_n)$$

for $n = 1,2,3,\dots$. In consequence, we derive by induction that

$$d(x_n,x_{n+1}) \le \theta^n d(x_0,x_1) \ , \quad \text{for} \quad n = 0,1,2,\dots \ .$$

Thus, if $n \ge 0$ is fixed and $m > n$, then

$$\begin{aligned} d(x_n,x_m) &\le d(x_n,x_{n+1}) + d(x_{n+1},x_{n+2}) + \cdots + d(x_{m-1},x_m) \\ &\le (\theta^n + \cdots + \theta^{m-1}) d(x_0,x_1) \\ &= \theta^n (1 + \cdots + \theta^{m-n-1}) d(x_0,x_1) \\ &= \theta^n \frac{1-\theta^{m-n}}{1-\theta} d(x_0,x_1) \\ &\le \frac{\theta^n}{1-\theta} d(x_0,x_1) \ . \end{aligned}$$

As $\theta < 1$, the right-hand side of the last inequality can be made as small as desired, independently of m, by taking n large enough. $\square$

Not only does this result provide existence and uniqueness, but the proof is constructive. Indeed, the proof consists of generating a sequence of approximations to $x = Tx$.

Corollary 8.9 (Fixed Point Iteration). *Suppose that (X, d) be a complete metric space, T a contraction mapping of X with contraction constant θ, and $x_0 \in X$. If the sequence $\{x_n\}_{n=0}^{\infty}$ is defined successively by $x_{n+1} = Tx_n$ for $n = 0, 1, 2, \ldots$, then $x_n \to x$, where x is the unique fixed point of T in X. Moreover,*

$$d(x_n, x) \leq \frac{\theta^n}{1-\theta} d(x_0, x_1) \ .$$

Example. Consider the initial value prooblem (IVP)

$$\begin{aligned} u_t &= \cos(u(t)) \ , \quad t > 0 \ , \\ u(0) &= u_0 \ . \end{aligned}$$

We would like to obtain a solution to the problem, at least up to some final time $T > 0$, using the fixed point theorem. At the outset we require two things: a complete metric space within which to seek a solution, and a map on that space for which a fixed point is the solution to our problem. It is not easy to handle the differential operator directly in this context, so we remove it through integration:

$$u(t) = u_0 + \int_0^t \cos(u(s))\, ds \ .$$

Now it is natural to seek a continuous function as a solution, say in $X = C^0([0, T])$, for some as yet unknown $T > 0$. It is also natural to consider the function

$$F(u) = u_0 + \int_0^t \cos(u(s))\, ds \ ,$$

which clearly takes X to X and has a fixed point at the solution to our IVP. To see if F is contractive, consider two functions u and v in X and compute

$$\begin{aligned} \|F(u) - F(v)\|_{L_\infty} &= \sup_{0 \leq t \leq T} \left| \int_0^t \big(\cos(u(s)) - \cos(v(s)) \big)\, ds \right| \\ &= \sup_{0 \leq t \leq T} \left| \int_0^t (-\sin(w(s))(u(s) - v(s))\, ds \right| \\ &\leq T \|u - v\|_{L_\infty} \ , \end{aligned}$$

wherein we have used the ordinary mean value theorem for functions of a real variable. So, if we take $T = 1/2$, we have a unique solution by the Banach Contraction Mapping Theorem. Since T is a fixed number independent of the solution u, we can iterate this process, starting at $t = 1/2$ (with "initial condition" $u(1/2)$) to extend the solution uniquely to $t = 1$, and so on, to obtain a solution for all time.

Example. Let $\kappa \in L_1(\mathbb{R})$, $\varphi \in C_B(\mathbb{R})$ and consider the nonlinear operator

$$\Phi u(x, t) = \varphi(x) + \int_0^t \int_{-\infty}^{\infty} \kappa(x - y)(u(y, s) + u^2(y, s))\, dy\, ds \ .$$

We claim that there exists $T = T(\|\varphi\|_\infty) > 0$ such that Φ has a fixed point in the space $X = C_B(\mathbb{R} \times [0, T])$.

Since κ is in $L_1(\mathbb{R})$, Φu makes sense. If $u \in C_B(\mathbb{R})$, then it is an easy exercise to see $\Phi u \in X$. Indeed, Φu is C^1 in the temporal variable and continuous in x by the Dominated Convergence Theorem.

Let $R > 0$ and B_R the closed ball of radius R about 0 in X. We want to show if R and T are chosen well, $\Phi : B_R \to B_R$ is a contraction. Let $u, v \in B_R$ and consider

$$\begin{aligned}
\|\Phi u - \Phi v\|_X &= \sup_{(x,t)\in\mathbb{R}\times[0,T]} \Big| \int_0^t \int_{-\infty}^{\infty} \kappa(x-y)(u - v + u^2 - v^2)\,dy\,ds \Big| \\
&\le T \sup_{(x,t)\in\mathbb{R}\times[0,T]} \int_{-\infty}^{\infty} |\kappa(x-y)(u - v + u^2 - v^2)|\,dy \\
&\le T\|\kappa\|_{L_1}\Big(\|u - v\|_X + \|u^2 - v^2\|_X\Big) \\
&\le T\|\kappa\|_{L_1}\Big(1 + \|u\|_X + \|v\|_X\Big)\|u - v\|_X \\
&\le T\|\kappa\|_{L_1}(1 + 2R)\|u - v\|_X \ .
\end{aligned}$$

Let

$$\theta = T(1 + 2R)\|\kappa\|_{L_1} \ ,$$

choose $R = 2\|\varphi\|_{L_\infty}$ and then choose T so that $\theta = 1/2$. With these choices, Φ is contractive on B_R and if $u \in B_R$, then indeed

$$\begin{aligned}
\|\Phi u\|_X &= \|\Phi u - \Phi 0\|_X + \|\Phi 0\|_X \\
&\le \theta\|u - 0\|_X + \|\varphi\|_{L_\infty} \\
&\le \frac{1}{2}R + \frac{1}{2}R = R \ .
\end{aligned}$$

That is, $\Phi : B_R \to B_R$, Φ is contractive, and B_R, being closed, is a complete metric space. We conclude that there exists a unique $u \in B_R$ such that

$$u = \Phi u \ .$$

Why do we care? Consider

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} + 2u\frac{\partial u}{\partial x} - \frac{\partial^3 u}{\partial x^2 \partial t} = 0 \ , \tag{8.22}$$

a nonlinear, dispersive wave equation. Write it as

$$(1 - \partial_x^2)u_t = -u_x - 2uu_x \equiv f \ .$$

The left-hand side is a nice operator, at least from the point of view of the Fourier Transform, as we will see in a moment, while the terms defining f are more troublesome. Take the Fourier transform on x to reach

$$(1 + \xi^2)\hat{u}_t = \hat{f} \ , \quad \text{i.e.,} \quad \hat{u}_t = \frac{1}{1+\xi^2}\hat{f} \ ,$$

whence, by taking the inverse Fourier transform, it is formally deduced that

$$u_t = \tilde{\kappa} * f = -\tilde{\kappa} * (u_x + 2uu_x) = -\tilde{\kappa} * (u + u^2)_x$$

where

$$\tilde{\kappa}(x) = \sqrt{2\pi}\mathcal{F}^{-1}\left(\frac{1}{1+\xi^2}\right) = \frac{1}{2}e^{-|x|} \ .$$

Let $\kappa = -\tilde{\kappa}_x \in L_1(\mathbb{R})$ to conclude

$$u_t(x,t) = \kappa * (u + u^2) \ .$$

Now integrate over $[0,t]$ and use the Fundamental Theorem of Calculus with constant of integration $u(x,0) = \varphi(x) \in C_B(\mathbb{R})$ to reach

$$u(x,t) = \varphi(x) + \int_0^t \kappa * (u + u^2)\, ds \ ,$$

which has the form with which we started the example. Thus our fixed point $\Phi u = u$ is formally a solution to (8.22), at least up to the time T, with the initial condition $u(x,0) = \varphi(x)$.

Corollary 8.10. *Let X be a Banach space and $f : X \to X$ a differentiable mapping. Suppose $\|Df(x)\|_{B(X,X)} \leq \kappa < 1$ for $x \in \overline{B_R(0)}$. If there is an $x_0 \in B_R(0)$ such that $\overline{B_r(x_0)} \subset B_R(0)$ for some $r \geq \|f(x_0) - x_0\|/(1-\kappa)$, then f has a fixed point in $B_R(0)$. Moreover, there is exactly one fixed point in $\overline{B_r(x_0)}$.*

That is, a map f which is locally contractive and for which we can find a point not moved too far by f has a fixed point, and the iteration

$$x_0,\ x_1 = f(x_0),\ x_2 = f(x_1), \cdots ,$$

generates a sequence that converges to a fixed point.

Proof. In fact, we show that f is a contraction mapping of $\overline{B_r(x_0)}$. First, by the Mean-Value Theorem, for any $x, y \in B_R(0)$,

$$\|f(x) - f(y)\| \leq \kappa \|x - y\| \ .$$

Hence f is contractive. The Contraction-Mapping Theorem will apply as soon as we know that f maps $\overline{B_r(x_0)}$ into itself, since $\overline{B_r(x_0)}$ is a complete metric space. By the triangle inequality, if $x \in \overline{B_r(x_0)}$,

$$\begin{aligned} \|f(x) - x_0\| &\leq \|f(x) - f(x_0)\| + \|f(x_0) - x_0\| \\ &\leq \kappa\|x - x_0\| + (1-\kappa) r \leq r \ . \qquad \square \end{aligned}$$

Theorem 8.11 (Simplified Newton Method). *Let X, Y be Banach spaces and $f : X \to Y$ a differentiable mapping. Suppose $A = Df(x_0)$ has a bounded inverse and that*

$$\|I - A^{-1} Df(x)\| \leq \kappa < 1 \tag{8.23}$$

for all $x \in B_r(x_0)$, for some $r > 0$. Let

$$\delta = \frac{(1-\kappa) r}{\|A^{-1}\|_{B(Y,X)}} \ .$$

Then the equation

$$f(x) = y$$

has a unique solution $x \in B_r(x_0)$ whenever $y \in B_\delta(f(x_0))$.

Proof. Let $y \in B_\delta(f(x_0))$ be given and define a mapping $g_y : X \to X$ by

$$g_y(x) = x - A^{-1}(f(x) - y) \ . \tag{8.24}$$

Notice that $g_y(x) = x$ if and only if $f(x) = y$. Note also that

$$Dg_y(x) = I - A^{-1} Df(x) \ ,$$

by the chain rule. By assumption, $\|Dg_y(x)\|_{B(X,X)} \leq \kappa < 1$ for $x \in B_r(x_0)$. Moreover, by the choice of y and δ,

$$\|g_y(x_0) - x_0\|_X = \|A^{-1}(f(x_0) - y)\|_X < (1-\kappa)r \ .$$

The hypotheses of Corollary 8.10 are verified, g_y is a contractive map of $\overline{B_r(x_0)}$, and the conclusion follows. $\square$

Remark. If $Df(x)$ is continuous as a function of x, then Hypothesis (8.23) is true for r small enough. Thus another conclusion is that at any point x where $Df(x)$ is boundedly invertible, there is an $r > 0$ and a $\delta > 0$ such that $f(B_r(x)) \supset B_\delta(f(x))$ and f is one-to-one on $B_r(x) \cap f^{-1}(B_\delta(f(x)))$. Hence there is a possibly smaller ball $B_t(x)$ such that f is one-to-one on $B_t(x)$ and $f(B_t(x)) \supset B_s(f(x))$ for some $s > 0$.

Notice the algorithm that is implied by the proof. Given y, start with a guess x_0 and form the sequence

$$x_{n+1} = g_y(x_n) = x_n - A^{-1}(f(x_n) - y) \ .$$

If things are as in the theorem, the sequence converges to the solution of $f(x) = y$ in $\overline{B_r(x_0)}$. Notice that if x is the solution, then

$$\begin{aligned} \|x_n - x\|_X &= \|g_y(x_{n-1}) - g_y(x)\|_X \\ &\leq \kappa \|x_{n-1} - x\|_X \\ &\leq \cdots \\ &\leq \kappa^n \|x_0 - x\|_X \ . \end{aligned}$$

More can be shown. We leave the rather lengthy proof of the following result to the reader.

Theorem 8.12 (Newton-Kantorovich Method). *Let X, Y be Banach spaces and $f : X \to Y$ a differentiable mapping. Assume that there is an $x_0 \in X$ and an $r > 0$ such that*

(i) *$A = Df(x_0)$ has a bounded inverse, and*
(ii) *$\|Df(x_1) - Df(x_2)\|_{B(X,Y)} \leq \kappa \|x_1 - x_2\|$*

for all $x_1, x_2 \in B_r(x_0)$. Let $y \in Y$ and set

$$\varepsilon = \|A^{-1}(f(x_0) - y)\|_X \ .$$

For any y such that

$$\varepsilon \leq \frac{r}{2} \quad \text{and} \quad 4\varepsilon\kappa \|A^{-1}\|_{B(Y,X)} \leq 1 \ ,$$

the equation

$$y = f(x)$$

has a unique solution in $B_r(x_0)$. Moreover, the solution is obtained as the limit of the Newton-iterates

$$x_{k+1} = x_k - Df(x_k)^{-1}(f(x_k) - y)$$

starting at x_0. The convergence is asymptotically quadratic; that is,

$$\|x_{k+1} - x_k\|_X \leq C \|x_k - x_{k-1}\|_X^2 \ ,$$

for k large, where C does not depend on k.

Theorem 8.13 (Inverse Function Theorem I). *Suppose the hypotheses of the Simplified Newton Method hold. Then the inverse mapping $f^{-1} : B_\delta(f(x_0)) \to B_r(x_0)$ is Lipschitz.*

PROOF. Let $y_1, y_2 \in B_\delta(f(x_0))$ and let x_1, x_2 be the unique points in $B_r(x_0)$ such that $f(x_i) = y_i$, for $i = 1, 2$. Fix a $y \in B_\delta(f(x_0))$, $y = y_0 = f(x_0)$ for example, and reconsider the mapping g_y defined in (8.24). As shown, g_y is a contraction mapping of $\overline{B_r(x_0)}$ into itself with Lipschitz constant $\kappa < 1$. Then

$$\begin{aligned}\|f^{-1}(y_1) - f^{-1}(y_2)\|_X &= \|x_1 - x_2\|_X \\ &= \|g_y(x_1) - g_y(x_2) + A^{-1}(f(x_2) - f(x_1))\|_X \\ &\leq \kappa\|x_1 - x_2\|_X + \|A^{-1}\|_{B(Y,X)}\|y_2 - y_1\|_Y \ .\end{aligned}$$

It follows that

$$\|x_1 - x_2\| \leq \frac{\|A^{-1}\|_{B(Y,X)}}{1-\kappa}\|y_1 - y_2\|$$

and hence that f^{-1} is Lipschitz with constant at most $\|A^{-1}\|_{B(Y,X)}/(1-\kappa)$. □

Earlier, we agreed that two Banach spaces X and Y are isomorphic if there is a $T \in B(X, Y)$ which is one-to-one and onto (and hence with bounded inverse by the Open Mapping Theorem). Isomorphic Banach spaces are indistinguishable as Banach spaces. A local version of this idea is now introduced.

DEFINITION. Let X, Y be Banach spaces and $U \subset X$, $V \subset Y$ open sets. Let $f : U \to V$ be one-to-one and onto. Then f is called a *diffeomorphism* on U and U is *diffeomorphic* to V if both f and f^{-1} are C^1, which is to say f and f^{-1} are Fréchet differentiable throughout U and V, respectively, and their derivatives are continuous on U and V, respectively. That is, the map

$$x \longmapsto Df(x) \ \text{ and } \ y \longmapsto Df^{-1}(y)$$

is continuous from U to $B(X, Y)$ and V to $B(Y, X)$, respectively.

THEOREM 8.14 (Inverse Function Theorem II). *Let X, Y be Banach spaces. Let $x_0 \in X$ be such that f is C^1 in a neighborhood of x_0 and $Df(x_0)$ is an isomorphism. Then there is an open set $U \subset X$ with $x_0 \in U$ and an open set $V \subset Y$ with $f(x_0) \in V$ such that $f : U \to V$ is a diffeomorphism. Moreover, for $y \in V$, $x \in U$, $y = f(x)$,*

$$D(f^{-1})(y) = (Df(x))^{-1} \ .$$

Before presenting the proof, we derive an interesting lemma. Let $GL(X, Y)$ denote the set of all isomorphisms of X onto Y. Of course, $GL(X, Y) \subset B(X, Y)$.

LEMMA 8.15. *Let X and Y be Banach spaces. Then $GL(X, Y)$ is an open subset of $B(X, Y)$. If $GL(X, Y) \neq \emptyset$, then the mapping $J_{X,Y} : GL(X, Y) \to GL(Y, X)$ given by $J_{X,Y}(A) = A^{-1}$ is one-to-one, onto, and continuous.*

PROOF. If $GL(X, Y) = \emptyset$, there is nothing to prove. Clearly $J_{Y,X}J_{X,Y} = I$ and $J_{X,Y}J_{Y,X} = I$, so $J_{X,Y}$ is both one-to-one and onto (but certainly *not* linear!). Let $A \in GL(X, Y)$ and $H \in B(X, Y)$. We claim that if $\|H\|_{B(X,Y)} < \theta/\|A^{-1}\|_{B(Y,X)}$ where $\theta < 1$, then $A + H \in GL(X, Y)$ also. To prove this, one need only show $A + H$ is one-to-one and onto.

We know that for any $|x| < 1$,

$$(1 + x)^{-1} = \sum_{n=0}^{\infty}(-x)^n \ ,$$

so consider the operators

$$S_N = A^{-1}\sum_{n=0}^{N}(-HA^{-1})^n \ , \qquad N = 1,2,\dots \ ,$$

in $B(Y,X)$. The sequence $\{S_N\}_{N=1}^{\infty}$ is Cauchy in $B(Y,X)$ since, for $M > N$,

$$\begin{aligned}
\|S_M - S_N\|_{B(Y,X)} &\le \|A^{-1}\|_{B(Y,X)} \sum_{n=N+1}^{M} \|(HA^{-1})^n\| \\
&\le \|A^{-1}\|_{B(Y,X)} \sum_{n=N+1}^{M} (\|H\|_{B(X,Y)}\|A^{-1}\|)^n_{B(Y,X)} \qquad (8.25)\\
&\le \|A^{-1}\|_{B(Y,X)} \sum_{n=N+1}^{M} \theta^n \to 0
\end{aligned}$$

as $N \to +\infty$. Hence $S_N \to S$ in $B(Y,X)$. Notice that

$$\begin{aligned}
(A+H)S &= \lim_{N\to\infty}(A+H)S_N \\
&= \lim_{N\to\infty}\sum_{n=0}^{N}(-HA^{-1})^n - \sum_{n=0}^{N}(-HA^{-1})^{n+1} \\
&= \lim_{N\to\infty}[I - (-HA^{-1})^{N+1}] \ .
\end{aligned}$$

But as $\|HA^{-1}\| \le \theta < 1$, $(HA^{-1})^N \to 0$ in $B(Y,Y)$. It is concluded that $(A+H)S = I$, and a similar calculation shows $S(A+H) = I$. Thus $A+H$ is one-to-one and onto, hence in $GL(X,Y)$. For use in a moment, notice that $\|S\|_{B(Y,X)} \le \|A^{-1}\|_{B(Y,X)}/(1-\theta)$, by an argument similar to (8.25).

For continuity, it suffices to take $A \in GL(X,Y)$ and show that $(A+H)^{-1} \to A^{-1}$ in $B(Y,X)$ as $H \to 0$ in $B(X,Y)$. But, as $S = (A+H)^{-1}$, this amounts to showing $S - A^{-1} \to 0$ as $H \to 0$. Now,

$$S - A^{-1} = (SA - I)A^{-1} = (S(A+H) - SH - I)A^{-1} = -SHA^{-1} \ .$$

Hence

$$\|S - A^{-1}\|_{B(Y,X)} \le \|S\|_{B(Y,X)}\|H\|_{B(X,Y)}\|A^{-1}\|_{B(Y,X)} \to 0$$

as $H \to 0$ since $\|A^{-1}\|_{B(Y,X)}$ is fixed and $\|S\|_{B(Y,X)} \le \|A^{-1}\|_{B(Y,X)}/(1-\theta)$ is bounded independently of H. $\square$

Proof of the Inverse Function Theorem II. Let $A = Df(x_0)$. Since f is a C^1-mapping, $Df(x) \to A$ in $B(X,Y)$ as $x \to x_0$ in X, so there is an $r' > 0$ such that

$$\|I - A^{-1}Df(x)\|_{B(X,X)} \le \frac{1}{2}$$

for all $x \in B_{r'}(x_0)$.

Because of Lemma 8.15, there is an r'' with $0 < r'' \le r'$ such that $Df(x)$ has a bounded inverse for all $x \in B_{r''}(x_0)$. It is further adduced that $Df(x)^{-1} \to A^{-1}$ as $x \to x_0$. In consequence, for $0 < r \le r''$, and for $x \in B_r(x_0)$,

$$\|Df(x)^{-1}\|_{B(Y,X)} \le 2\|A^{-1}\|_{B(Y,X)} \ .$$

Appealing now to the Simplified Newton Method, it is concluded that there is an $r > 0$ and a $\delta > 0$ such that $f : U \to V$ is one-to-one, and onto, where

$$V = B_\delta(f(x_0)) \quad \text{with} \quad \delta = \frac{r}{2\|A^{-1}\|_{B(Y,X)}}$$

and

$$U = B_r(x_0) \cap f^{-1}(V) \ .$$

It remains to establish that f^{-1} is a C^1 mapping with the indicated derivative. Suppose it is known that

$$Df^{-1}(y) = Df(x)^{-1} \ , \quad \text{when} \quad y = f(x) \ , \tag{8.26}$$

where $x \in U$ and $y \in V$. In this case, the mapping from y to $Df^{-1}(y)$ is obtained in three steps, namely

$$y \longmapsto f^{-1}(y) \longmapsto Df(f^{-1}(y)) \longmapsto Df(f^{-1}(y))^{-1} = Df^{-1}(y) \ ,$$
$$Y \xrightarrow{f^{-1}} X \xrightarrow{Df} B(X,Y) \xrightarrow{J} B(Y,X) \ .$$

As all three of these components is continuous, so is the composite.

Thus it is only necessary to establish (8.26). To this end, fix $y \in V$ and let k be small enough that $y + k$ also lies in V. If $x = f^{-1}(y)$ and $h = f^{-1}(y+k) - x$, then

$$\begin{aligned} \|f^{-1}(y+k) - f^{-1}(y) - Df(x)^{-1}k\|_X &= \|h - Df(x)^{-1}[f(x+h) - f(x)]\|_X \\ &= \|Df(x)^{-1}[f(x+h) - f(x) - Df(x)h]\|_X \\ &\le 2\|A^{-1}\|_{B(Y,X)}\|f(x+h) - f(x) - Df(x)h\|_Y \ . \end{aligned} \tag{8.27}$$

The right-hand side of (8.27) tends to 0 as $h \to 0$ in X since f is differentiable at x. Hence if we show that $h \to 0$ as $k \to 0$, it follows that f^{-1} is differentiable at $y = f(x)$ and that

$$Df^{-1}(y) = Df(x)^{-1} \ .$$

The theorem is thereby established because of our earlier remarks. But,

$$\|h\|_X = \|f^{-1}(y+k) - f^{-1}(y)\|_X \le M\|k\|_Y \ ,$$

since f^{-1} is Lipschitz (see Theorem 8.13). $\square$

THEOREM 8.16 (Implicit Function Theorem). *Let X, Y, Z be Banach spaces and suppose*

$$f : Z \times X \to Y$$

to be a C^1-mapping defined at least in a neighborhood of a point (z_0, x_0). Denote by y_0 the image $f(z_0, x_0)$. Suppose $D_x f(z_0, x_0) \in GL(X, Y)$. Then there are open sets

$$W \subset Z \ , \qquad U \subset X \ , \qquad V \subset Y$$

with $z_0 \in W$, $x_0 \in U$ and $y_0 \in V$ and a unique mapping

$$g : W \times V \to U$$

such that

$$f(z, g(z,y)) = y \tag{8.28}$$

for all $(z, y) \in W \times V$. Moreover, g is C^1 and

$$Dg(z,y)(\eta, \zeta) = D_x f(z,x)^{-1}(\zeta - D_z f(z,x)\eta)$$

for $(z, y) \in W \times V$ *and* $(\eta, \zeta) \in Z \times Y$*.*

REMARK. If $Z = \{0\}$ is the trivial Banach space, this result recovers the Inverse Function Theorem.

PROOF. Define an auxiliary mapping $\hat{f}$ by

$$\hat{f}(z, x) = (z, f(z, x)) \ .$$

Then $f : Z \times X \to Z \times Y$ and f is C^1 since both its components are. Moreover, from Proposition 8.7 it is adduced that

$$D\hat{f}(z, x)(\eta, \varphi) = (\eta, D_z f(z, x)\eta + D_x f(z, x)\varphi)$$

for (z, x) in the domain of f and $(\eta, \varphi) \in Z \times X$. If $D_x f(z, x)$ is an invertible element of $B(X, Y)$, then $D\hat{f}$ is an invertible element of $B(Z \times X, Z \times Y)$ and its inverse is given by

$$D\hat{f}(z, x)^{-1}(\eta, \zeta) = (\eta, D_x f(z, x)^{-1}(\zeta - D_z f(z, x)\eta) \ ,$$

as one checks immediately. The Inverse Function Theorem implies $\hat{f}$ is a diffeomorphism from some open set $\hat{U}$ about (z_0, x_0) to an open set $\hat{V}$ containing (z_0, y_0). By continuity of the projections onto components in $Z \times X$, there are open sets W and V in Z and Y, respectively, such that $W \times V \subset \hat{V}$. By construction

$$\hat{f}^{-1}(z, y) = (z, g(z, y))$$

where g is a C^1-mapping. And, since

$$(z, y) = \hat{f}(\hat{f}^{-1}(z, y)) = \hat{f}(z, g(z, y)) = (z, f(z, g(z, y))) \ ,$$

g solves the equation (8.28). $\square$

COROLLARY 8.17. *Let f be as in Theorem 8.16. Then there is a unique C^1-branch of solutions of the equation*

$$f(z, y) = y_0$$

defined in a neighborhood of (z_0, x_0)*.*

PROOF. Let $h(z) = g(z, y_0)$ in the Implicit Function Theorem. Then h is C^1, $h(z_0) = x_0$, and

$$f(z, h(z)) = y_0$$

for z near z_0. $\square$

EXAMPLE. The eigenvalues of an $n \times n$ matrix are given as the roots of the characterictis polynomial

$$p(A, \lambda) = \det(A - \lambda I) \ .$$

In fact, p is a polynomial in λ and all entries of A, so it is C^1 as a function $p : \mathbb{C}^{n\times n} \times \mathbb{C} \to \mathbb{C}$. Fix A_0 and λ_0 such that λ_0 is a simple (i.e., nonrepeated) root of A_0. Then $D_2 p(A_0, \lambda_0) \neq 0$ (i.e., $D_2 p(A_0, \lambda_0) \in GL(\mathbb{C}, \mathbb{C})$), so every matrix A near A_0 has a unique eigenvalue

$$\lambda = g(A, 0) = \hat{g}(A) \ ,$$

where $\hat{g}$ is C^1. As we change A continuously from A_0, the eigenvalue λ_0 changes continuously until possibly it becomes a repeated eigenvalue, at which point a bifurcation may occur. A bifurcation cannot occur otherwise.

Example. Consider the ordinary differential initial value problem

$$u' = 1 - u + \epsilon e^u \ , \quad 0 < t \ ,$$
$$u(0) = 0 \ .$$

If $\epsilon = 0$, this is a well posed linear problem with solution

$$u_0(t) = 1 - e^{-t}$$

which exists for all time t. It is natural to consider if there is a solution for $\epsilon > 0$. Note that if ϵ is very large, then we have essentially the equation

$$w' = \epsilon e^w \ ,$$

which has solution

$$w(t) = -\log(1 - \epsilon t) \to \infty \text{ as } t \to 1/\epsilon \ .$$

Thus we do not have a solution w for all time. The Implicit Function Theorem clarifies the situation. Our parameter space is $Z = \mathbb{R}$, and our function space is $X = \{f \in C^1(0, \infty) : f(0) = 0\}$. We have a mapping $T : Z \times X \to Y = C^0(0, \infty)$ defined by

$$T(\epsilon, u) = u' - 1 + u - \epsilon e^u \ ,$$

which is C^1; in fact, the partial derivatives are

$$D_Z T(\epsilon, u)(z, v) = z e^u \quad \text{and} \quad D_X T(\epsilon, u)(z, v) = v' + v - \epsilon v e^u \ .$$

Now $D_X T(0, u)(z, v) = v' + v$ maps one-to-one and onto, since we can uniquely solve $v' + v = f$ by using an integrating factor. Thus the Implicit Function Theorem gives us an $\epsilon_0 >$ such that for $|\epsilon| < \epsilon_0$, there exists a solution defined for all time. Moreover, there is a unique solution in a neighborhood of u_0 in X.

8.3. Higher Derivatives

Here, consideration is given to higher-order Fréchet derivatives. The development starts with some helpful preliminaries.

Definition. Let X, Y be vector spaces over $\mathbb{F}$. A n-linear map is a function

$$f : \underbrace{X \times \cdots \times X}_{n\text{-components}} \longrightarrow Y$$

for which f is linear in each argument separately. The set of all n-linear maps from X to Y is denoted $\mathcal{B}^n(X, Y)$. By convention, we take $\mathcal{B}^0(X, Y) = Y$.

Proposition 8.18. *Let X, Y be NLS's and let $n \in \mathbb{N}$. The following are equivalent for $f \in \mathcal{B}^n(X, Y)$.*

(i) *f is continuous,*
(ii) *f is continuous at 0,*
(iii) *f is bounded, which is to say there is a constant M such that*

$$\|f(x_1, \dots, x_n)\|_Y \le M \|x_1\|_X \cdots \|x_n\|_X \ .$$

We denote by $B^n(X, Y)$ the subspace of $\mathcal{B}^n(X, Y)$ of all bounded n-linear maps, and we let $B^0(X, Y) = Y$. Moreover, $B^1(X, Y) = B(X, Y)$.

PROPOSITION 8.19. *Let X, Y be NLS's and $n \in \mathbb{N}$. For $f \in B^n(X, Y)$, define*

$$\|f\|_{B^n(X,Y)} = \sup_{\substack{x_i \in X: \\ \|x_i\| \le 1 \\ 1 \le i \le n}} \|f(x_1, \dots, x_n)\|_Y \ .$$

Then $\|\cdot\|_{B^n(X,Y)}$ is a norm on $B^n(X, Y)$ and if Y is complete, so is $B^n(X, Y)$.

PROPOSITION 8.20. *Let k, ℓ be non-negative integers and X, Y NLS's. Then $B^k(X, B^\ell(X, Y))$ is isomorphic to $B^{k+\ell}(X, Y)$ and the norms are the same.*

PROOF. Let $n = k + \ell$ and define $J : B^k(X, B^\ell(X, Y)) \to B^n(X, Y)$ by

$$(Jf)(x_1, \dots, x_n) = f(x_1, \dots, x_k)(x_{k+1}, \dots, x_n) \ .$$

This makes sense because $f(x_1, \dots, x_k) \in B^\ell(X, Y)$. Clearly $Jf \in \mathcal{B}^n(X, Y)$, and

$$\begin{aligned} \|Jf\|_{B^n(X,Y)} &= \sup_{\substack{\|x_i\| \le 1 \\ 1 \le i \le n}} \|Jf(x_1, \dots, x_n)\|_Y \\ &= \sup_{\substack{\|x_i\| \le 1 \\ 1 \le i \le k}} \|f(x_1, \dots, x_k)\|_{B^\ell(X,Y)} \\ &= \|f\|_{B^k(X, B^\ell(X,Y))} \ , \end{aligned}$$

so $Jf \in B^n(X, Y)$. For $g \in B^n(X, Y)$, define $\hat{g} \in \mathcal{B}^k(X, \mathcal{B}^\ell(X, Y))$ by

$$\hat{g}(x_1, \dots, x_k)(x_{k+1}, \dots, x_n) = g(x_1, \dots, x_n) \ .$$

A straightforward calculation shows that

$$\|\hat{g}\|_{B^k(X, B^\ell(X,Y))} \le \|g\|_{B^n(X,Y)} \ ,$$

so $\hat{g} \in B^k(X, B^\ell(X, Y))$ and $J\hat{g} = g$. Thus J is a one-to-one, onto, bounded linear map, so it also has a bounded inverse and is in fact an isomorphism. Moreover, J is norm-preserving by the above bounds (i.e., $\|Jf\|_{B^n(X,Y)} = \|f\|_{B^k(X, B^\ell(X,Y))}$). $\square$

DEFINITION. Let X, Y be Banach spaces and $f : X \to Y$. For $n = 2, 3, \dots$, define f to be n-times Fréchet differentiable in a neighborhood of a point x if f is $(n-1)$-times differentiable in a neighborhood of x and the mapping $x \mapsto D^{n-1} f(x)$ is Fréchet differentiable near x. Define

$$D^n f(x) = D D^{n-1} f(x) \ , \qquad n = 2, 3, \dots \ .$$

Notice that

$$\begin{gathered} f : X \to Y \ , \\ Df : X \to B(X, Y) \ , \\ D^2 f = D(Df) : X \to B(X, B(X, Y)) = B^2(X, Y) \ , \\ \vdots \\ D^n f = D(D^{n-1} f) : X \to B(X, B^{n-1}(X, Y)) = B^n(X, Y) \ . \end{gathered}$$

EXAMPLES. 1. If $A \in B(X, Y)$, then $DA(x) = A$ for all x. Hence

$$D^2 A(x) \equiv 0 \quad \text{for all} \quad x \ .$$

This is because

$$DA(x + h) - DA(x) \equiv 0$$

for all x.

2. Let $X = H$ be a Hilbert space, $\mathbb{F} = \mathbb{R}$, and $A \in B(H, H)$. Define $f : H \to \mathbb{R}$ by

$$f(x) = (x, Ax)_H \ .$$

Then, $Df(x) = \mathcal{R}((A+A^*)x)$, where $\mathcal{R}$ denotes the Riesz map. That is, $Df(x) \in B(H, \mathbb{R}) = H^*$, and for $y \in H$,

$$Df(x)(y) = (y, A^*x + Ax)_H \ .$$

To compute the second derivative, form the difference

$$[Df(x+h) - Df(x)]y = (y, (A+A^*)(x+h) - (A+A^*)x) = (y, (A+A^*)h) \ ,$$

for $y \in H$. Thus it is determined that

$$D^2f(x)(y, h) = (y, (A+A^*)h) \ .$$

Note that $D^2f(x)$ does not depend on x, so $D^3f(x) \equiv 0$.

3. Let $K \in L_\infty(I \times I)$ where $I = [a, b] \subset \mathbb{R}$. Define $F : L_p(I) \to L_p(I)$ by

$$F(g)(x) = \int_I K(x, y) g^p(y)\, dy$$

for $p \in \mathbb{N}$ and $x \in I$. Then, $DF(g) \in B(L_p(I), L_p(I))$ and

$$DF(g)h = p \int_I K(x, y) g^{p-1}(y) h(y)\, dy \ ,$$

since the Binomial Theorem gives the expansion

$$\begin{aligned} F(g+h) - F(g) &= \int_I K(x, y)[(g+h)^p - g^p]\, dy \\ &= \int_I K(x, y) \left[p g^{p-1}(y) h(y) + \binom{p}{2} g^{p-2}(y) h^2(y) + \cdots \right] dy \ , \end{aligned}$$

wherein all but the first term is higher-order in h. Thus it follows readily that

$$\begin{aligned} DF(g+h)u - DF(g)u &= p \int_I K(x, y) \left[(g+h)^{p-1}u - g^{p-1}u \right] dy \\ &= p(p-1) \int_I K(x, y)[g^{p-2}hu]\, dy + \text{ terms cubic in } \ h, u \ . \end{aligned}$$

It follows formally, and can be verified under strict hypotheses, that

$$D^2F(g)(h, k) = p(p-1) \int_I K(x, y) g^{p-2}(y) h(y) k(y)\, dy \ .$$

Lemma 8.21 (Schwarz). *Let X, Y be Banach spaces, U an open subset of X and $f : U \to Y$ have two derivatives. Then $D^2f(x)$ is a symmetric bilinear mapping.*

Proof. Consider the difference

$$g(h, k) = f(x+h+k) - f(x+h) - f(x+k) + f(x) - D^2f(x)(h, k) \ ,$$

so that

$$\begin{aligned} \|D^2f(x)(h, k) - D^2f(x)(k, h)\|_Y &= \|g(h, k) - g(k, h)\|_Y \\ &\le \|g(h, k) - g(0, k)\|_Y + \|g(0, k) - g(k, h)\|_Y \\ &= \|g(h, k) - g(0, k)\|_Y + \|g(0, h) - g(k, h)\|_Y \end{aligned}$$

since $g(0,k) = g(0,h) = 0$. But the right-hand side of the last equality is bounded above by the Mean Value Theorem as

$$\|g(h,k) - g(0,k)\|_Y \le \sup \|D_1 g\|_{B(X,Y)} \|h\|_X \ ,$$
$$\|g(k,h) - g(0,h)\|_Y \le \sup \|D_1 g\|_{B(X,Y)} \|k\|_X \ .$$

Differentiate g partially with respect to the first variable h to obtain

$$\begin{aligned} D_1 g(h,k)\tilde{h} &= Df(x+h+k)\tilde{h} - Df(x+h)\tilde{h} - D^2 f(x)(\tilde{h},k) \\ &= Df(x+h+k)\tilde{h} - Df(x)\tilde{h} - D^2 f(x)(\tilde{h}, h+k) \\ &\quad - [Df(x+h)\tilde{h} - Df(x)\tilde{h} - D^2 f(x)(\tilde{h},h)] \ . \end{aligned}$$

For $\|h\|_X$, $\|k\|_X$ small, it follows from the definition of the Fréchet derivative of Df that

$$\|D_1 g(h,k)\|_{B(X,Y)} = o(\|h\|_X + \|k\|_X) \ .$$

Thus we have established that

$$\|D^2 f(x)(h,k) - D^2 f(x)(k,h)\|_Y = o(\|k\|_X + \|h\|_X)\,(\|h\|_X + \|k\|_X) \ ,$$

and it follows from bilinearity that in fact

$$D^2 f(x)(h,k) = D^2 f(x)(k,h)$$

for all $h,k \in X$. □

Corollary 8.22. *Let f, X, Y and U be as in Lemma 8.21, but suppose f has $n \ge 2$ derivatives in U. Then $D^n f(x)$ is symmetric under permutation of its arguments. That is, if π is an $n \times n$ symmetric permutation matrix, then*

$$D^n f(x)(h_1, \dots, h_n) = D^n f(x)(\pi(h_1, \dots, h_n)) \ .$$

Proof. This follows by induction from the fact that $D^n f(x) = D^2(D^{n-2} f)(x)$. □

Theorem 8.23 (Taylor's Formula). *Let X, Y be Banach spaces, $U \subset X$ open and suppose $f : U \to Y$ has n derivatives throughout U. Then for h small,*

$$f(x+h) = f(x) + Df(x)h + \frac{1}{2}D^2 f(x)(h,h) + \dots + \frac{1}{n!}D^n f(x)(h,\dots,h) + R_n(x,h) \tag{8.29}$$

and

$$\frac{\|R_n(x,h)\|_Y}{\|h\|_X^n} \longrightarrow 0$$

as $h \to 0$ in X, i.e., $\|R_n(x,h)\|_Y = o(\|h\|_X^n)$.

Proof. We first note in general that if $F \in B^m(X,Y)$ is symmetric and g is defined by

$$g(h) = F(h, \dots, h) \ ,$$

then

$$Dg(h)k = mF(h, \dots, h, k) \ .$$

This follows by straightforward calculation. For $m = 1$, F is just a linear map and the result is already known. For $m = 2$, for example, just compute

$$g(h+k) - g(h) - 2F(h,k) = F(h+k, h+k) - F(h,h) - 2F(h,k) = F(k,k) \ ,$$

and

$$\|F(k,k)\|_Y \le C\|k\|_X^2 \ ,$$

showing g is differentiable and that $Dg(h) = 2F(h,\cdot)$.

For the theorem, the case $n = 1$ just reproduces the definition of f being differentiable at x. We initiate an induction on n, supposing the result valid for all functions f satisfying the hypotheses for $k < n$, where $n \geq 2$. Let f satisfy the hypotheses for $k = n$. Define R_n as in (8.29) and notice that

$$D_2R_n(x,h) = Df(x+h) - Df(x) - D^2f(x)(h,\cdot) - \cdots - \frac{1}{(n-1)!}D^nf(x)(h,\dots,h,\cdot)\ .$$

That is,

$$Df(x+h) = Df(x) + D^2f(x)(h,\cdot) + \cdots + \frac{1}{(n-1)!}D^nf(x)(h,\dots,h,\cdot) + D_2R_n(x,h)\ ,$$

which is the $(n-1)$st Taylor expansion of Df, and by induction we conclude that

$$\frac{\|D_2R_n(x,h)\|_{B(X,Y)}}{\|h\|_X^{n-1}} \longrightarrow 0$$

as $h \to 0$. On the other hand, by the Mean-Value Theorem, if $\|h\|_X$ is sufficiently small, then

$$\frac{\|R_n(x,h)\|_Y}{\|h\|_X^n} = \frac{\|R_n(x,h) - R_n(x,0)\|_Y}{\|h\|_X^n} \leq \sup_{0\leq\alpha\leq1} \frac{\|D_2R_n(x,\alpha h)\|_{B(X,Y)}}{\|h\|_X^{n-1}} \longrightarrow 0$$

as $h \to 0$. $\square$

8.4. Extrema

DEFINITION. Let X be a set and $f : X \to \mathbb{R}$. A point $x_0 \in X$ is a *minimum* if $f(x_0) \leq f(x)$ for all $x \in X$; it is a *maximum* if $f(x_0) \geq f(x)$ for all $x \in X$. An *extrema* is a point which is a maximum or a minimum. If X has a topology, we say x_0 is a *relative* (or *local*) minimum if there is an open set $U \subset X$ with $x_0 \in U$ such that

$$f(x_0) \leq f(x)$$

for all $x \in U$. Similarly, if

$$f(x_0) \geq f(x)$$

for all $x \in U$, then x_0 is a *relative maximum*. If equality is disallowed above when $x \neq x_0$, the (relative) minimum or maximum is said to be *strict*.

THEOREM 8.24. *Let X be a NLS, let U be an open set in X and let $f : U \to \mathbb{R}$ be differentiable. If $x_0 \in U$ is a relative maximum or minimum, then $Df(x_0) = 0$.*

PROOF. We show the theorem when x_0 is a relative minimum; the other case is similar. We argue by contradiction, so suppose that $Df(x_0)$ is not the zero map. Then there is some $h \neq 0$ such that $Df(x_0)h \neq 0$. By possibly reversing the sign of h, we may assume that $Df(x_0)h < 0$. Let $t_0 > 0$ be small enough that $x_0 + th \in U$ for $|t| \leq t_0$ and consider for such t

$$\begin{aligned}\frac{1}{t}\left[f(x_0+th) - f(x_0)\right] &= \frac{1}{t}\left[Df(x_0)(th) + R_1(x_0,th)\right]\\ &= Df(x_0)h + \frac{1}{t}R_1(x_0,th)\ .\end{aligned}$$

The quantity $R_1(x_0,th)/t \to 0$ as $t \to 0$. Hence for $t_1 \leq t_0$ small enough and $|t| \leq t_1$,

$$\left|\frac{1}{t}R_1(x_0,th)\right| \leq \frac{1}{2}|Df(x_0)h|\ .$$

It follows that for $|t| < t_1$,

$$f(x_0 + th) = f(x_0) + t\left[Df(x_0)h + \frac{1}{t}R_1(x_0, th)\right] < f(x_0)\ ,$$

provided we choose $t > 0$. This contradiction proves the result for relative minima. Similar ruminations establish the conclusion for relative maxima. □

DEFINITION. A *critical point* of a mapping $f : U \to Y$, where U is open in X, is a point x_0 where $Df(x_0) = 0$. This is also referred to as a *stationary point* by some authors.

COROLLARY 8.25. *If $f : U \to \mathbb{R}$ is differentiable, then the relative extrema of f in U are critical points of f.*

DEFINITION. Let X be a vector space over $\mathbb{R}$, $U \subset X$ a convex subset, and $f : U \to \mathbb{R}$. We say that f is *convex* if whenever $x_1, x_2 \in U$ and $\lambda \in (0, 1)$, then

$$f(\lambda x_1 + (1 - \lambda)x_2) \le \lambda f(x_1) + (1 - \lambda)f(x_2)\ .$$

We say that f is *concave* if the opposite inequality holds. Moreover, we say that f is *strictly* convex or concave if equality is not allowed above.

PROPOSITION 8.26. *Linear functionals on X are both convex and concave (but not strictly so). If $a, b > 0$ and f, g are convex, then $af + bg$ is convex, and if at least one of f or g is strictly convex, then so is $af + bg$. Furthermore, f is (strictly) convex if and only if $-f$ is (strictly) concave.*

We leave the proof as an easy exercise of the definitions.

PROPOSITION 8.27. *Let X be a NLS, U a convex subset of X, and $f : U \to \mathbb{R}$ convex and differentiable. Then, for $x, y \in U$,*

$$f(y) \ge f(x) + Df(x)(y - x)\ ,$$

and, if $Df(x) = 0$, then x is a minimum of f in U. Moreover, if f is strictly convex, then for $x \ne y$,

$$f(y) > f(x) + Df(x)(y - x)\ ,$$

and $Df(x) = 0$ implies that f has a strict and therefore unique minimum.

PROOF. By convexity, for $\lambda \in [0, 1]$,

$$\lambda f(y) + (1 - \lambda)f(x) \ge f(x + \lambda(y - x))\ ,$$

whence

$$f(y) - f(x) \ge \frac{f(x + \lambda(y - x)) - f(x)}{\lambda}\ .$$

Take the limit as $\lambda \to 0$ on the right-hand side to obtain the desired result.

We leave the proof of the strictly convex case to the reader. □

EXAMPLE. Let $\Omega \subset \mathbb{R}^d$, $f \in L_2(\Omega)$, and assume that the underlying field is real. Define $J : H_0^1(\Omega) \to \mathbb{R}$ by

$$J(v) = \tfrac{1}{2}\|\nabla v\|^2_{L_2(\Omega)} - (f, v)_{L_2(\Omega)}\ .$$

We claim that $\|\nabla v\|^2_{L_2(\Omega)}$ is strictly convex. To verify this, let $v, w \in H^1_0(\Omega)$ and $\lambda \in (0,1)$. Then

$$\begin{aligned}
&\|\nabla(\lambda v + (1-\lambda)w)\|^2_{L_2(\Omega)} \\
&\quad = \lambda^2 \|\nabla v\|^2_{L_2(\Omega)} + (1-\lambda)^2 \|\nabla w\|^2_{L_2(\Omega)} + 2\lambda(1-\lambda)\,(\nabla v, \nabla w)_{L_2(\Omega)} \\
&\quad = \lambda \|\nabla v\|^2_{L_2(\Omega)} + (1-\lambda)\|\nabla w\|^2_{L_2(\Omega)} - \lambda(1-\lambda)[\|\nabla v\|^2_{L_2(\Omega)} + \|\nabla w\|^2_{L_2(\Omega)} - 2(\nabla v, \nabla w)_{L_2(\Omega)}] \\
&\quad = \lambda \|\nabla v\|^2_{L_2(\Omega)} + (1-\lambda)\|\nabla w\|^2_{L_2(\Omega)} - \lambda(1-\lambda)\,(\nabla(v-w), \nabla(v-w))_{L_2(\Omega)} \\
&\quad < \lambda \|\nabla v\|^2_{L_2(\Omega)} + (1-\lambda)\|\nabla w\|^2_{L_2(\Omega)} \ ,
\end{aligned}$$

unless $v - w$ is identically constant on each connected component of Ω. As $v - w \in H^1_0(\Omega)$, $v = w$ on $\partial\Omega$, and so $v = w$ everywhere. That is, we have strict inequality whenever $v \neq w$, and so we conclude that $\|v\|^2_{L_2(\Omega)}$ is strictly convex. By Prop. 8.26, we conclude that $J(v)$ is also strictly convex. Moreover,

$$DJ(u,v) = (\nabla u, \nabla v)_{L_2(\Omega)} - (f,v)_{L_2(\Omega)} \ .$$

We conclude that $u \in H^1_0(\Omega)$ satisfies the boundary value problem

$$(\nabla u, \nabla v)_{L_2(\Omega)} = (f,v)_{L_2(\Omega)}$$

if and only if u minimizes the "energy functional" $J(v)$ over $H^1_0(\Omega)$:

$$J(u) < J(v) \quad \text{for all } v \in H^1_0(\Omega), v \neq u \ .$$

Moreover, such a function u is unique.

Local convexity suffices to verify that a critical point is a relative extrema. More generally, we can examine the second derivative.

Theorem 8.28. *If X is a NLS and $f : X \to \mathbb{R}$ is twice differentiable at a relative minimum $x \in X$, then*

$$D^2 f(x)(h,h) \geq 0 \quad \textit{for all } h \in X \ .$$

Proof. By Taylor's formula

$$f(x \pm \lambda h) = f(x) \pm Df(x)\lambda h + \tfrac{1}{2}\lambda^2 D^2 f(x)(h,h) + o(\lambda^2 \|h\|^2_X) \ ,$$

so we conclude that

$$D^2 f(x)(h,h) = \lim_{\lambda \to 0} \frac{f(x+\lambda h) + f(x-\lambda h) - 2f(x)}{\lambda^2} \geq 0$$

if x is a local minimum. □

Remark. In infinite dimensions, it is *not* the case that $Df(x) = 0$ and $D^2 f(x)(h,h) > 0$ for all $h \neq 0$ implies that x is a local minimum. For example consider the function $f : \ell_2 \to \mathbb{R}$ defined by

$$f(x) = \sum_{k=1}^{\infty} \left(\frac{1}{k} - x_k \right) x_k^2 \ ,$$

where $x = (x_k)_{k=1}^\infty \in \ell_2$. Note that f is well defined on ℓ_2 (i.e., the sum converges). Direct calculation shows that

$$Df(x)(h) = \sum_{k=1}^\infty \left(\frac{2}{k} - 3x_k\right) x_k h_k \ ,$$

$$D^2 f(x)(h,h) = \sum_{k=1}^\infty \left(\frac{2}{k} - 6x_k\right) h_k^2 \ ,$$

so $f(0) = 0$, $Df(0) = 0$, and $D^2 f(0)(h,h) > 0$ for all $h \neq 0$. However, let x^k be the element of ℓ_2 such that x_j^k is 0 if $j \neq k$ and $2/k$ if $j = k$. We compute that $f(x^k) < 0$, in spite of the fact that $x^k \to 0$ as $k \to \infty$. Thus 0 is not a local minimum of f.

THEOREM 8.29 (Second Derivative Test). *Let X be a NLS, and $f : X \to \mathbb{R}$ have two derivatives at a critical point $x \in X$. If there is some constant $c > 0$ such that*

$$D^2 f(x)(h,h) \geq c\|h\|_X^2 \quad \text{for all } h \in X \ ,$$

then x is a strict local minimum point.

PROOF. By Taylor's Theorem, for any $\varepsilon > 0$, there is $\delta > 0$ such that for $\|h\|_X \leq \delta$,

$$|f(x+h) - f(x) - \tfrac{1}{2} D^2 f(x)(h,h)| \leq \varepsilon \|h\|_X^2 \ ,$$

since the Taylor remainder is $o(\|h\|_X^2)$. Thus,

$$f(x+h) - f(x) \geq \tfrac{1}{2} D^2 f(x)(h,h) - \varepsilon\|h\|_X^2 \geq (\tfrac{1}{2}c - \varepsilon)\|h\|_X^2 \ ,$$

and taking $\varepsilon = c/4$, we conclude that

$$f(x+h) \geq f(x) + \tfrac{1}{4} c\|h\|_X^2 \ ,$$

i.e., f has a local minimum at x. □

REMARK. This theorem is not as general as it appears. If we define the bilinear form

$$(h,k)_X = D^2 f(x)(h,k) \ ,$$

we easily verify that, with the assumption of the Second Derivative Test, that in fact $(h,k)_X$ is an inner product, which induces a norm equivalent to the original. Thus in fact X must be a pre-Hilbert space, and it makes no sense to attempt use of the theorem when X is known not to be pre-Hilbert.

8.5. The Euler-Lagrange Equations

A common problem in science and engineering applications is to find extrema of a functional that involves an integral of a function. We will consider this situation via the following problem. Let $a < b$,

$$f : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

and define the functional $F : C^1([a,b]) \to \mathbb{R}$ by

$$F(y) = \int_a^b f(x, y(x), y'(x))\, dx \ .$$

With α and β given in $\mathbb{R}^n$, let

$$C^1_{\alpha,\beta}([a,b],\mathbb{R}^n) = \{v : [a,b] \to \mathbb{R}^n \mid v \text{ has a continuous first derivative,} \\ v(a) = \alpha, \text{ and } v(b) = \beta\} \ .$$

Our goal is to find $y \in C^1_{\alpha,\beta}([a,b],\mathbb{R}^n)$ such that

$$F(y) = \min_{v \in C^1_{\alpha,\beta}([a,b],\mathbb{R}^n)} F(v) \ .$$

EXAMPLE. Find $y(x) \in C^1([a,b])$ such that $y(a) = \alpha$ and $y(b) = \beta$ and the surface of revolution of the graph of y about the x-axis has minimal area. Recall that a differential of arc length is given by

$$ds = \sqrt{1 + (y'(x))^2}\, dx \ ,$$

so our area as a function of the curve y is

$$A(y) = \int_a^b 2\pi y(x) \sqrt{1 + (y'(x))^2}\, dx \ . \tag{8.30}$$

If α and β are zero, $C^1_{0,0}([a,b],\mathbb{R}^n) = C^1_0([a,b],\mathbb{R}^n)$ is a Banach space with the $W^{1,1}([a,b])$ (Sobolev space) maximum norm, and our minimum is found at a critical point. However, in general $C^1_{\alpha,\beta}([a,b],\mathbb{R}^n)$ is not a linear vector space. Rather it is an affine space, a translate of a vector space. To see this, let

$$\ell(x) = \frac{1}{b-a}[\alpha(b-x) + \beta(x-a)]$$

be the linear function connecting (a,α) to (b,β). Then

$$C^1_{\alpha,\beta}([a,b],\mathbb{R}^n) = C^1_0([a,b],\mathbb{R}^n) + \ell \ .$$

To solve our problem, then, we need to consider any fixed element of $C^1_{\alpha,\beta}$, such as $\ell(x)$, and all possible "admissible variations" h of it that lie in C^1_0; that is, we minimize $F(v)$ by searching among all possible "competing functions" $v = \ell + h \in C^1_{\alpha,\beta}$, where $h \in C^1_0$, for the one that minimizes $F(v)$, if any. On C^1_0, we can find the derivative of $F(\ell + h)$ as a function of h, and thereby restrict our search to the critical points. We call such a point $y = \ell + h$ a *critical point* for F defined on $C^1_{\alpha,\beta}$. We present a general result on the derivative of F of the form considered in this section.

THEOREM 8.30. *If $f \in C^1([a,b] \times \mathbb{R}^n \times \mathbb{R}^n)$ and*

$$F(y) = \int_a^b f(x, y(x), y'(x))\, dx \ ,$$

then $F : C^1([a,b]) \to \mathbb{R}$ is continuously differentiable and

$$DF(y)(h) = \int_a^b [D_2 f(x, y(x), y'(x))\, h(x) + D_3 f(x, y(x), y'(x))\, h'(x)]\, dx$$

for all $h \in C^1([a,b])$.

PROOF. Let A be defined by

$$Ah = \int_a^b [D_2 f(x, y(x), y'(x))\, h(x) + D_3 f(x, y(x), y'(x))\, h'(x)]\, dx \ ,$$

which is clearly a bounded linear functional on C^1, since the norm of any $v \in C^1$ is

$$\|v\| = \max(\|v\|_{L_\infty}, \|v'\|_{L_\infty}) \ .$$

Now

$$\begin{aligned} F(y+h) - F(y) &= \int_a^b \int_0^1 \frac{d}{dt} f(x, y+th, y'+th')\, dt\, dx \\ &= \int_a^b \int_0^1 D_2 f(x, y+th, y'+th')\, h + D_3 f(x, y+th, y'+th')\, h'\, dt\, dx \ , \end{aligned}$$

so

$$\begin{aligned} |F(y+h) - F(y) - Ah| \le \int_a^b \int_0^1 &|[D_2 f(x, y+th, y'+th') - D_2 f(x,y,y')]\, h|\, dt\, dx \\ + \int_a^b \int_0^1 &|[D_3 f(x, y+th, y'+th') - D_3 f(x,y,y')]\, h'|\, dt\, dx \ . \end{aligned}$$

Since $D_2 f$ and $D_3 f$ are uniformly continuous on compact sets, the right-hand side is $o(\|h\|)$, and we conclude that $DF(y) = A$.

It remains to show that $DF(y)$ is continuous. But this follows from uniform continuity of $D_2 f$ and $D_3 f$, and from the computation

$$\begin{aligned} |DF(y+h)k - DF(y)k| \le \int_a^b &|[D_2 f(x, y+h, y'+h') - D_2 f(x,y,y')]\, k|\, dx \\ + \int_a^b &|[D_3 f(x, y+h, y'+h') + D_3 f(x,y,y')]\, k'|\, dx \ , \end{aligned}$$

which tends to 0 as $\|h\| \to 0$ for any $k \in C^1([a,b])$ with $\|k\| \le 1$. □

THEOREM 8.31. *Suppose $f \in C^1([a,b] \times \mathbb{R}^n \times \mathbb{R}^n)$, $y \in C^1_{\alpha,\beta}([a,b])$, and*

$$F(y) = \int_a^b f(x, y(x), y'(x))\, dx \ .$$

Then y is a critical point for F if and only if the curve $x \mapsto D_3 f(x, y(x), y'(x))$ is $C^1([a,b])$ and y satisfies the Euler-Lagrange Equations

$$D_2 f(x,y,y') - \frac{d}{dx} D_3 f(x,y,y') = 0 \ .$$

In component form,the Euler-Lagrange Equations are

$$\frac{\partial f}{\partial y_k} = \frac{d}{dx} \frac{\partial f}{\partial y'_k}, \quad k = 1, ..., n \ ,$$

or

$$f_{y_k} = \frac{d}{dx} f_{y'_k}, \quad k = 1, ..., n \ .$$

The converse implication of the Theorem is easily shown from the previous result after integrating by parts, since $h \in C^1_0$. The direct implication follows easily from the previous result and the following Lemma. We leave the details to the reader.

LEMMA 8.32 (Dubois-Reymond). *Let φ and ψ lie in $C^0([a,b], \mathbb{R}^n)$. Then*

(i) $\int_a^b \varphi(x) \cdot h'(x)\, dx = 0$ *for all $h \in C^1_0$ if and only if φ is identically constant.*

(ii) $\int_a^b [\varphi(x) \cdot h(x) + \psi(x) \cdot h'(x)]\, dx = 0$ *for all $h \in C^1_0$ if and only if $\psi \in C^1$ and $\psi' = \varphi$.*

PROOF. Both converse implications are trivial after integrating by parts. For the direct implication of (i), let

$$\bar{\varphi} = \frac{1}{b-a}\int_a^b \varphi(x)\,dx \ ,$$

and note that then

$$0 = \int_a^b \varphi(x)\cdot h'(x)\,dx = \int_a^b (\varphi(x) - \bar{\varphi})\cdot h'(x)\,dx \ .$$

Take

$$h = \int_a^x (\varphi(s) - \bar{\varphi})\,ds \in C_0^1 \ ,$$

so that $h' = \varphi - \bar{\varphi}$. We thereby demonstrate that

$$\|\varphi - \bar{\varphi}\|_{L_2} = 0 \ ,$$

and conclude that $\varphi = \bar{\varphi}$ (almost everywhere, but both functions are continuous, so everywhere).

For the direct implication of (ii), let

$$\Phi = \int_a^x \varphi(s)\,ds \ ,$$

so that $\Phi' = \varphi$. Then the hypothesis of (ii) shows that

$$\int_a^b [\Phi - \psi]\cdot h'\,dx = \int_a^b [\Phi\cdot h'(x) + \varphi\cdot h]\,dx = \int_a^b \frac{d}{dx}(\Phi\cdot h)\,dx = 0 \ ,$$

since h vanishes at a and b. We conclude from (i) that $\Phi - \psi$ is constant. Since Φ is C^1, so is ψ, and $\psi' = \Phi' = \varphi$. $\square$

DEFINITION. Solutions of the Euler-Lagrange equations are called *extremals.*

EXAMPLE. We illustrate the theory by finding the shortest path between two points. Suppose $y(x)$ is a path in $C^1_{\alpha,\beta}([a,b])$, which connects (a,α) to (b,β). Then we seek to minimize the length functional

$$L(y) = \int_a^b \sqrt{1 + (y'(x))^2}\,dx$$

over all such y. The integrand is

$$f(t,y,y') = \sqrt{1 + (y'(x))^2} \ ,$$

so the Euler-Lagrange equations become simply

$$(D_3 f)' = 0 \ ,$$

and so we conclude that for some constant c,

$$(1 + (y'(x))^2)(y'(x))^2 = c^2 \ .$$

Thus,

$$y'(x) = \pm\sqrt{\frac{c^2}{1-c^2}} \ ,$$

if $c^2 \neq 1$, and there is no solution otherwise. In any case, $y'(x)$ is constant, so the only critical paths are lines, and there is a unique such line in $C^1_{\alpha,\beta}([a,b])$. Since $L(y)$ is convex, this path is necessarily a minimum, and we conclude the well-known maxim: the shortest distance between two points is a straight line.

EXAMPLE. Many problems have no solutions. For example, consider the problem of minimizing the length of the curve $y \in C^1([0,1])$ such that $y(0) = y(1) = 0$ and $y'(0) = 1$. The minimum approaches 1, but is never attained by a C^1-function.

It is generally not easy to solve the Euler-Lagrange equations. They constitute a nonlinear second order ordinary differential equation for $y(x)$. To see this, suppose that $y \in C^2([a,b])$ and compute

$$D_2 f = (D_3 f)' = D_1 D_3 f + D_2 D_3 f\, y' + D_3^2 f\, y'' \ ,$$

or, provided $D_3^2 f(x, y, y')$ is invertible,

$$y'' = (D_3^2 f)^{-1}(D_2 f - D_1 D_3 f - D_2 D_3 f\, y') \ .$$

DEFINITION. If y is an extremal and $D_3^2 f(x, y, y')$ is invertible for all $x \in [a,b]$, then we call y a *regular* extremal.

PROPOSITION 8.33. *If $f \in C^2([a,b] \times \mathbb{R}^n \times \mathbb{R}^n)$ and $y \in C^1([a,b])$ is a regular extremal, then $y \in C^2([a,b])$.*

In this case, we can reduce the problem to first order.

THEOREM 8.34. *If $f \in C^2([a,b] \times \mathbb{R}^n \times \mathbb{R}^n)$, $f(x,y,z) = f(y,z)$ only, and $y \in C^1([a,b])$ is a regular extremal, then $y'\, D_3 f - f$ is constant.*

PROOF. Simply compute

$$\begin{aligned} (y'\, D_3 f - f)' &= y''\, D_3 f + y'(D_3 f)' - f' \\ &= y''\, D_3 f + y'\, D_2 f - (D_2 f\, y' + D_3 f\, y'') = 0 \ , \end{aligned}$$

using the Euler-Lagrange equation for the extremal. ☐

EXAMPLE. We reconsider the problem of finding $y(x) \in C^1([a,b])$ such that $y(a) = \alpha$ and $y(b) = \beta$ and the surface of revolution of the graph of y about the x-axis has minimal area. The area as a function of the curve is given in (8.30), so

$$f(y, y') = 2\pi y(x)\sqrt{1 + (y'(x))^2} \ .$$

Note that

$$D_3 f(y, y') = \frac{2\pi y}{(1 + (y')^2)^{3/2}} \neq 0 \ ,$$

unless $y = 0$. Thus our nonzero extremals are regular, so we can use the theorem to find them. For some constant C,

$$\frac{2\pi y\, (y')^2}{(1 + (y')^2)^{1/2}} - 2\pi y(1 + (y')^2)^{1/2} = 2\pi C \ ,$$

which implies that

$$y' = \frac{1}{C}\sqrt{y^2 - C^2} \ .$$

Applying separation of variables, we need to integrate

$$\frac{dy}{\sqrt{y^2 - C^2}} = \frac{dx}{C} \ ,$$

which, for some constant λ, gives us the solution

$$y(x) = C \cosh(x/C + \lambda) ,$$

which is called a *catenary*. Suppose that $a = 0$, so that $C = \alpha/\cosh\lambda$ and

$$y(b) = \beta = \frac{\alpha}{\cosh\lambda} \cosh\left(\frac{\cosh\lambda}{\alpha} b + \lambda\right) .$$

That is, we determine C once we have λ, which must solve the above equation. There may or may not be solutions λ (i.e., there may not be regular extremals). It is a fact, which we will not prove (see [**Sa**, pp. 62ff.]), that the minimal area is given either by a regular extremal or the Goldschmidt solution, which is the piecewise graph that uses straight lines to connect the points $(0,\alpha)$ to $(0,0)$, $(0,0)$ to $(b,0)$, and finally $(b,0)$ to (b,β). This is not a C^1 curve, so it is technically inadmissible, but it has area $A_G = \pi(\alpha^2+\beta^2)$. If there are no extremals, then, given $\epsilon > 0$, we have C^1 curves approximating the Goldschmidt solution such that the area is greater than but within ϵ of A_G.

Example (The Brachistochrone problem with a free end). Sometimes one does not impose a condition at one end. An example is the Brachistochrone problem. Consider a particle moving under the influence of gravity in the xy-plane, where y points upwards. We assume that the particle starts from rest at the position $(0,0)$ and slides frictionlessly along a curve $y(x)$, moving in the x-direction a distance $b > 0$ and falling an unspecified distance (see Fig. 1). We wish to minimize the total travel time. Let the final position be (b,β), where $\beta < 0$ is unspecified. We assume that the curve

$$y \in C^1_*([0,b]) = \{v \in C^1([0,b]) : v(0) = 0\} .$$

The steeper the curve, the faster it will move; however, it must convert some of this speed into motion in the x-direction to travel distance b. To derive the travel time functional $T(y)$, we note that Newton's Law implies that for a mass m traveling on the arc s with angle θ from the downward direction (see Fig. 1),

$$m\frac{d^2 s}{dt^2} = mg\cos\theta = mg\frac{dy}{ds} ,$$

where g is the gravitational constant. The mass cancels and

$$\frac{1}{2}\frac{d}{dt}\left(\frac{ds}{dt}\right)^2 = \frac{d^2 s}{dt^2}\frac{ds}{dt} = g\frac{dy}{dt} ,$$

so we conclude that for some constant C,

$$\left(\frac{ds}{dt}\right)^2 = 2gy + C .$$

But at $t = 0$, both the speed and $y(0)$ are zero, so $C = 0$, and

$$\frac{ds}{dt} = \sqrt{2gy} .$$

Now the travel time is given by

$$T(y) = \int dt = \int \frac{ds}{\sqrt{2gy}} = \int_0^b \sqrt{\frac{1+(y'(x))^2}{2gy}}\, dx .$$

We need a general result to deal with the free end.

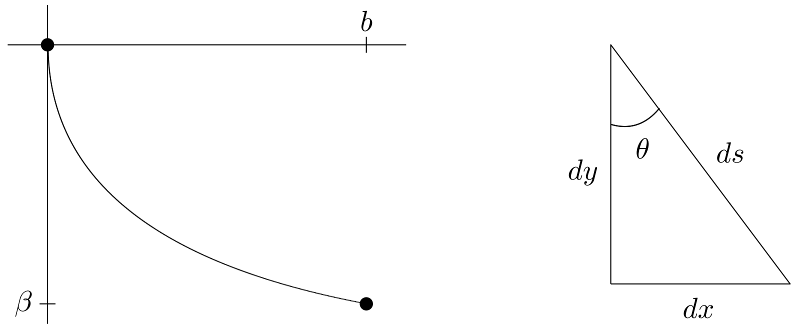

FIGURE 1. The Brachistochrone problem.

THEOREM 8.35. *If $y \in C^2([a,b])$ minimizes*

$$F(y) = \int_a^b f(x, y(x), y'(x))\, dx$$

subject only to the single constraint that $y(a) = \alpha \in \mathbb{R}$, then y must satisfy the Euler Lagrange equations and $D_3 f(b, y(b), y'(b)) = 0$.

PROOF. We simply compute for $y \in C^1_*([a,b]) + \alpha$ and $h \in C^1_*([a,b])$

$$DF(y)\, h = \int_a^b (D_2 f\, h + D_3 f\, h')\, dx = \int_a^b (D_2 f\, h - (D_3 f)'\, h)\, dx + D_3 f(b, y(b), y'(b))\, h(b) \ .$$

If $h \in C^1_{0,0}([a,b])$, we derive the Euler-Lagrange equations, and otherwise we obtain the second condition at $x = b$. □

EXAMPLE (The Brachistochrone problem with a free end, continued). Since we are looking for a minimum, we can drop the factor $\sqrt{2g}$ and concentrate on

$$f(y, y') = \sqrt{\frac{1 + (y'(x))^2}{y}} \ .$$

This is independent of x, so we solve

$$y' D_3 f - f = C_1 = \frac{1}{\sqrt{y}} \left(\frac{(y')^2}{\sqrt{1 + (y'(x))^2}} + \sqrt{1 + (y'(x))^2} \right) ,$$

or

$$\int \sqrt{\frac{y}{C_1^{-2} - y}}\, dy = x - C_2 \ .$$

This is solved using a trigonometric substitution, so we let

$$y = C_1^{-2} \sin(\phi/2) = (1 - \cos\phi)/2C^2 \ ,$$

and then

$$x = (\phi - \sin\phi)/2C^2 + C_2 \ .$$

Applying the initial condition ($\phi = 0$), we determine that the curve is

$$(x, y) = C(\phi - \sin\phi, 1 - \cos\phi)$$

for some constant C. This is a *cycloid*. Now C is determined by the auxiliary condition

$$0 = D_3 f(y(b), y'(b)) = \frac{1}{\sqrt{y(b)}} \frac{y'(b)}{\sqrt{1 + (y'(b))^2}} \ ,$$

which requires

$$0 = y'(b) = \frac{dy}{d\phi}\left(\frac{dx}{d\phi}\right)^{-1} = \frac{\sin\phi(b)}{1 - \cos\phi(b)} \ .$$

Thus $\phi(b) = \pi$ (since $\phi \in [0, \pi]$), so $C = b/\pi$ and the solution is complete.

8.6. Constrained Extrema and Lagrange Multipliers

When discussing the Euler-Lagrange equations, we considered the problem of finding relative extrema of a nonlinear functional in $C^1_{\alpha,\beta}$, which is an affine translate of a Banach space. We can phrase this differently: we found extrema in the Banach space C^1 subject to the linear constraint that the function agrees with α and β at its endpoints. We consider now the more general problem of finding relative extrema of a nonlinear functional subject to a possibly nonlinear constraint.

Let X be a Banach space, $U \subset X$ open, and $f : U \to \mathbb{R}$. To describe our constraint, we assume that there are functions $g_i : X \to \mathbb{R}$ for $i = 1, ..., m$ that define the set $M \subset U$ by

$$M = \{x \in U : g_i(x) = 0 \text{ for all } i\} \ .$$

Our problem is to find the relative extrema of f restricted to M. Note that M is not necessarily open, so we must discuss what happens on ∂M. To rephrase our problem: Find the relative extrema of $f(x)$ on U subject to the constraints

$$g_1(x) = ... = g_m(x) = 0 \ . \tag{8.31}$$

Example. Consider a thin membrane stretched over a rigid frame. We describe this as follows. Let $\Omega \in \mathbb{R}^2$ be open in the xy-plane and suppose that there is a function $f : \partial\Omega \to \mathbb{R}$ which describes the z-coordinate (height) of the rigid frame. That is, the frame is

$$\{(x, y, z) : z = f(x, y) \text{ for all } (x, y) \in \partial\Omega\} \ .$$

We let $u : \Omega \to \mathbb{R}$ be the height of the membrane. The membrane will assume that shape that minimizes the energy, subject to the constraint that it attaches to the rigid frame. The energy functional $E : H^1(\Omega) \to \mathbb{R}$ is a sum of the elastic energy and the gravitational potential energy:

$$E(u) = \int_\Omega [\tfrac{1}{2} c|\nabla u|^2 + gu] \, dx \ ,$$

where c is a constant related to the elasticity of the membrane and g is the gravitational constant. We minimize E subject to the constraint that the trace of u, $\gamma_0(u)$, agrees with f on the boundary, that is, that $G(u) = 0$, where $G : H^1(\Omega) \to \mathbb{R}$ is defined by

$$G(u) = \int_{\partial\Omega} |\gamma_0(u) - f| \, ds \ .$$

To find the relative extrema of $f(x)$ on U subject to the constraints (8.31), we can instead solve un unconstrained problem, albeit in more dimensions. Define $H : X \times \mathbb{R}^m \to \mathbb{R}$ by

$$H(x, \lambda) = f(x) + \lambda_1 g_1(x) + ... + \lambda_m g_m(x) \ . \tag{8.32}$$

The critical points of H are given by solving for a root of the system of equations defined by the partial derivatives

$$
\begin{aligned}
&D_1H(x,\lambda) = Df(x) + \lambda_1 Dg_1(x) + ... + \lambda_m Dg_m(x)\ ,\\
&D_2H(x,\lambda) = g_1(x)\ ,\\
&\vdots\\
&D_{m+1}H(x,\lambda) = g_m(x)\ .
\end{aligned}
$$

Such a critical point satisfies the m constraints and an additional condition which is necessary for an extrema, as we now prove.

THEOREM 8.36 (Lagrange Multiplier Theorem). *Let X be a Banach space, $U \subset X$ open, and $f, g_i : U \to \mathbb{R}$, $i = 1, ..., m$, be continuously differentiable. If $x \in M$ is a relative extrema for $f|_M$, where*

$$M = \{x \in U : g_i(x) = 0 \text{ for all } i\}\ ,$$

then there is a nonzero $\lambda = (\lambda_0, ..., \lambda_m) \in \mathbb{R}^{m+1}$ such that

$$\lambda_0 Df(x) + \lambda_1 Dg_1(x) + ... + \lambda_m Dg_m(x) = 0\ . \tag{8.33}$$

That is, to find a local extrema in M, we need only consider points that satisfy (8.33). We search through the unconstrained space U for such points x, and then we must verify that in fact $x \in M$ holds. Two possibilities arise for $x \in U$. If $\{Dg_i(x)\}_{i=1}^m$ is linearly independent, the only nontrivial way to satisfy (8.33) is to take $\lambda_0 \neq 0$. Otherwise, $\{Dg_i(x)\}_{i=1}^m$ is linearly dependent, and (8.33) is satisfied for a nonzero λ with $\lambda_0 = 0$.

Our method of search then is clear. (1) First we find critical points of H as defined above in (8.32). These points automatically satisfy both (8.33) and $x \in M$. These points are potential relative extrema. (2) Second, we find points $x \in U$ where $\{Dg_i(x)\}_{i=1}^m$ is linearly dependent. Then (8.33) is satisfied, so we must further check to see if indeed $x \in M$, i.e., each $g_i(x) = 0$. If so, x is also a potential relative extrema. (3) Finally, we determine if the potential relative extrema are indeed extrema or not. Often, the constraints are chosen so that $\{Dg_i(x)\}_{i=1}^m$ is always linearly independent, and the second step does not arise. (We remark that if we want extrema on $\bar{M}$, then we would also need to chect points on ∂M.)

PROOF OF THE LAGRANGE MULTIPLIER THEOREM. Suppose that x is a local minimum of $f|_M$; the case of a local maximum is similar. Then we can find an open set $V \subset U$ such that $x \in V$ and

$$f(x) \le f(y) \quad \text{for all } y \in M \cap V\ .$$

Define $F : V \to \mathbb{R}^{m+1}$ by

$$F(y) = (f(y), g_1(y), ..., g_m(y))\ .$$

Since x is a local minimum on M, for any $\varepsilon > 0$,

$$(f(x) - \varepsilon, 0, ..., 0) \neq F(y) \quad \text{for all } y \in V\ .$$

Thus, we conclude that F does *not* map V onto an open neighborhood of $F(x) = (f(x), 0, ..., 0) \in \mathbb{R}^{m+1}$.

Suppose that $DF(x)$ maps X onto $\mathbb{R}^{m+1}$. Then construct a space $\tilde{X} = \text{span}\{v_1, ..., v_{m+1}\} \subset X$ where we choose each v_i such that $DF(x)(v_i) = e_i$, the standard unit vector in the ith direction in $\mathbb{R}^{m+1}$. Let $\hat{X} = \{v \in \tilde{X} : x + v \in V\}$, and define the function $h : \hat{X} \to \mathbb{R}^{m+1}$ by

$h(v) = F(x+v)$. Now $Dh(0) = DF(x)$ maps $\tilde{X}$ onto $\mathbb{R}^{m+1}$ is invertible, so the Inverse Function Theorem implies that h maps an open subset S of $\hat{X}$ containing 0 onto an open subset of $\mathbb{R}^{m+1}$ containing $h(0) = F(x)$. But then $x + S \subset V$ is an open set that contradicts our previous conclusion regarding F.

Thus $DF(x)$ cannot map onto all of $\mathbb{R}^{m+1}$, and so it maps onto a proper subspace. There then is some nonzero vector $\lambda \in \mathbb{R}^{m+1}$ orthogonal to $DF(x)(X)$. Thus

$$\lambda_0 Df(x)(y) + \lambda_1 Dg_1(x)(y) + ... + \lambda_m Dg_m(x)(y) = 0 ,$$

for any $y \in X$, and we conclude that this linear conbination of the operators must vanish, i.e., (8.33) holds. □

Example. The Isoperimetric Problem can be stated as follows: among all rectifiable curves in $\mathbb{R}^2_+$ from $(-1,0)$ to $(1,0)$ with length ℓ, find the one enclosing the greatest area. We need to maximize the functional

$$A(u) = \int_{-1}^{1} u(t)\,dt$$

subject to the constraint

$$L(u') = \int_{-1}^{1} \sqrt{1+(u'(t))^2}\,dt = \ell$$

over the set $u \in C^1_{0,0}([-1,1])$ with $u \geq 0$. Let

$$H(u,\lambda) = A(u) + \lambda[L(u') - \ell] = \int_{-1}^{1} h_\lambda(u,u')\,dt ,$$

where

$$h_\lambda(u,u') = u + \lambda\sqrt{1+(u'(t))^2} - \ell/2 .$$

To find a critical point of the system, we need to find both $D_u H$ and $D_\lambda H$. For the former, it is given by considering λ fixed and solving the Euler-Lagrange equations: $D_2 h_\lambda = (D_3 h_\lambda)'$. That is,

$$1 = \lambda \frac{d}{dt}\left(\frac{u'}{\sqrt{1+(u'(t))^2}}\right) ,$$

so for some constant C_1,

$$t = \lambda \frac{u'}{\sqrt{1+(u'(t))^2}} + C_1 .$$

Solving for u' yields

$$u'(t) = \frac{t - C_1}{\sqrt{\lambda^2 - (t-C_1)^2}} .$$

Another integration gives a constant C_2 and

$$u(t) = \sqrt{\lambda^2 - (t-C_1)^2} + C_2 ,$$

or, rearranging, we obtain the equation of a circular arc

$$(u(t) - C_2)^2 + (t - C_1)^2 = \lambda^2$$

with center (C_1, C_2) of radius λ. The partial derivative $D_\lambda H$ simply recovers the constraint that the arc length is ℓ, and the requirement that $u \in C_{0,0}([-1,1])$ says that it must go through the points $u(-1) = (-1,0)$ and $u(1) = (1,0)$. We leave it to the reader to complete the example by showing that these conditions uniquely determine $C_1 = 0$, C_2, and $\lambda = \sqrt{1+C_2^2}$, where C_2 satisfies the transcendental equation

$$\sqrt{1+C_2^2} = \frac{\ell}{2[\pi - \tan^{-1}(1/C_2)]} \ .$$

Moreover, the reader may justify that a maximum is obtained at this critical point.

We also need to check the condition $DL(u') = 0$. Again the Euler-Lagrange equations allow us to find these points easily. The result, left to the reader, is that for some constant C of integration,

$$u' = \frac{C}{\sqrt{1-C}} \ ,$$

which means that u is a straight line. The fixed ends imply that $u \equiv 0$, and so we do not satisfy the length constraint unless $\ell = 2$, a trivial case to analyze.

As a corollary, among curves of fixed lengths, the circle encloses the region of greatest area.

8.7. Lower Semi-Continuity and Existence of Minima

Whether there exists a minimum of a functional is an important question. If a minimum exists, we can locate it by analyzing critical points. Perhaps the simplest criterion for the existence of a minimum is to consider convex functionals, as we have done previously. Next simplest is perhaps to note that a continuous function on a compact set attains its minimum.

However, in an infinite dimensional Banach space X, bounded sets are not compact; that is, compact sets are very small. This observation suggests that, at least when X is reflexive, we consider using the weak topology, since then the Banach-Alaoglu Theorem 2.31 implies that bounded sets are weakly compact. The problem now is that many interesting functionals are not weakly continuous, such as the norm itself. For the norm, it is easily seen that:

$$\text{If } u_n \overset{w}{\rightharpoonup} u, \text{ then } \liminf_{n\to\infty} \|u_n\| \geq \|u\| \ ,$$

with inequality possible. We are lead to consider a weaker notion of continuity.

Definition. Let X be a topological space. A function $f : X \to (-\infty, \infty]$ is said to be *lower semicontinuous* (l.s.c.) if whenever $\lim_{n\to\infty} x_n = x$, then

$$\liminf_{n\to\infty} f(x_n) \geq f(x) \ .$$

Proposition 8.37. *Let X be a topological space and $f : X \to (-\infty,\infty]$. Then f is lower semicontinuous if and only if the sets*

$$A_\alpha = \{x \in X : f(x) \leq \alpha\}$$

are closed for all $\alpha \in \mathbb{R}$.

Proof. Suppose f is l.s.c. Let $x_n \in A_\alpha$ be such that $x_n \to x \in X$. Then

$$f(x) \leq \liminf_{n\to\infty} f(x_n) \leq \alpha \ ,$$

so $x \in A_\alpha$ and A_α is closed.

Suppose now each A_α is closed. Then

$$A_\alpha^c = \{x \in X : f(x) > \alpha\}$$

is open. Let $x_n \to x \in X$, and suppose that $x \in A_\alpha^c$ for some α (i.e., $f(x) > \alpha$). Then there is some $N_\alpha > 0$ such that for all $n \geq N_\alpha$, $x_n \in A_\alpha^c$, and so $\liminf_{n\to\infty} f(x_n) \geq \alpha$. In other words, whenever $f(x) > \alpha$, $\liminf_{n\to\infty} f(x_n) \geq \alpha$, so we conclude that

$$\liminf_{n\to\infty} f(x_n) \geq \sup\{\alpha : f(x) > \alpha\} = f(x) \ .$$

□

THEOREM 8.38. *If M is compact and $f : M \to (-\infty, \infty]$ is lower semicontinuous, then f is bounded below and takes on its minumum value.*

PROOF. Let

$$A = \inf_{x\in M} f(x) \in [-\infty, \infty] \ .$$

If $A = -\infty$, choose a sequence $x_n \in M$ such that $f(x_n) \leq -n$ for all $n \geq 1$. Since M is compact, there is $x \in M$ such that, for some subsequence, $x_{n_i} \to x$ as $i \to \infty$. But

$$f(x) \leq \liminf_{i\to\infty} f(x_{n_i}) = -\infty \ ,$$

contradicting that f maps into $(-\infty, \infty]$. Thus $A > -\infty$, and f is bounded below.

Now choose a sequence $x_n \in M$ such that $f(x_n) \leq A + 1/n$, and again extract a convergent subsequence $x_{n_i} \to x \in M$ as $i \to \infty$. We compute

$$A \leq f(x) \leq \liminf_{i\to\infty} f(x_{n_i}) \leq \liminf_{i\to\infty}(A + 1/n) = A \ ,$$

and we conclude that $f(x) = A$ attains its minimum at x. □

The previous results apply to general topological spaces. For reflexive Banach spaces, we have both the strong (or norm) and weak topologies.

THEOREM 8.39. *Let M be a weakly closed subspace of a reflexive Banach space X. If $f : M \to (-\infty, \infty]$ is weakly lower semicontinuous and, for some α, $A_\alpha = \{x \in X : f(x) \leq \alpha\}$ is bounded and nonempty, then f is bounded from below and there is some $x_0 \in M$ such that*

$$f(x_0) = \min_{x\in M} f(x) \ .$$

PROOF. By the Banach-Alaoglu Theorem 2.31, $\bar{A}_\alpha$ is compact, so $f|_{\bar{A}_\alpha}$ attains its minimum. But for $x \in M \setminus \bar{A}_\alpha$, $f(x) > \alpha \geq \min_{x\in\bar{A}_\alpha} f(x)$, and the theorem follows. □

It is important to determine when a function is weakly lower semicontinuous. The following requirement is left to the reader, and its near converse follows.

PROPOSITION 8.40. *If X is a Banach space and $f : X \to (-\infty, \infty]$ is weakly lower semicontinuous, then f is strongly lower semicontinuous.*

THEOREM 8.41. *Suppose X is a Banach space and $f : X \to (-\infty, \infty]$. If $V = \{x \in X : f(x) < \infty\}$ is a subspace of X, and if f is both convex on V and strongly lower semicontinuous, then f is weakly lower semicontinuous.*

PROOF. For $\alpha \in \mathbb{R}$, let $A_\alpha = \{x \in X : f(x) \leq \alpha\}$ be as usual. Since f is strongly l.s.c., Prop. 8.37 implies that A_α is closed in the strong (i.e., norm) topology. But f being convex on V implies that A_α is also convex. A strongly closed convex set is weakly closed (see Corollary 2.37), so we conclude that f is weakly l.s.c. □

LEMMA 8.42. *Let $f : \mathbb{C} \to [0,\infty)$ be convex, Ω a domain in $\mathbb{R}^d$, and $1 \le p < \infty$. Then $F : L_p(\Omega) \to [0,\infty]$, defined by*

$$F(u) = \int_\Omega f(u(x))\,dx \ ,$$

is norm and weak l.s.c.

PROOF. Since F is convex, it is enough to prove the norm l.s.c. property. Let $u_n \to u$ in $L_p(\Omega)$ and choose a subsequence such that

$$\lim_{i\to\infty} F(u_{n_i}) = \liminf_{n\to\infty} F(u_n)$$

and $u_{n_i}(x) \to u(x)$ for almost every $x \in \Omega$. Then $f(u_{n_i}(x)) \to f(u(x))$ for a.e. x, since f being convex is also continuous. Fatou's lemma finally implies that

$$F(u) \le \liminf_{i\to\infty} F(u_{n_i}) = \liminf_{n\to\infty} F(u_n) \ .$$

□

COROLLARY 8.43. *If Ω is a domain in $\mathbb{R}^d$ and $1 \le p, q < \infty$, then the $L_q(\Omega)$-norm is weakly l.s.c. on $L_p(\Omega)$.*

We close this section with two examples that illustrate the concepts.

EXAMPLE. Let $f \in C_0^\infty(\mathbb{R}^n)$ and consider the differential equation

$$-\Delta u + u|u| + u = f \ .$$

Let us show that there is a solution. Let

$$F(u) = \int_{\mathbb{R}^d} \left(\tfrac{1}{2}|\nabla u|^2 + \tfrac{1}{3}|u|^3 + \tfrac{1}{2}|u|^2 - fu\right) dx \ ,$$

which may be $+\infty$ for some u. Now if $v \in C_0^\infty(\mathbb{R}^d)$,

$$\begin{aligned} DF(u)(v) &= \int_{\mathbb{R}^d} \big(\nabla u \cdot \nabla v + |u|uv + uv - fv\big)\,dx \\ &= \int_{\mathbb{R}^d} \big(-\Delta u + u|u| + u - f\big)\,v\,dx \end{aligned}$$

which vanishes if and only if the differential equation is satisfied. Since F is clearly convex, there will be a solution to the differential equation if F takes on its minimum.

Now

$$F(u) \ge \tfrac{1}{2}\|\nabla u\|^2_{L_2(\mathbb{R}^d)} - \|f\|_{L_2(\mathbb{R}^d)}\|u\|_{L_2(\mathbb{R}^d)} \ge \tfrac{1}{4}\|\nabla u\|^2_{L_2(\mathbb{R}^d)} - \|f\|^2_{L_2(\mathbb{R}^d)} \ ,$$

so the set $\{u \in L_2(\mathbb{R}^d) : F(u) \le 1\}$ is bounded by $4(1 + \|f\|^2_{L_2(\mathbb{R}^d)})$, and nonempty (since it contains $u \equiv 0$). We will complete the proof if we can show that F is l.s.c.

The last term of F is weakly continuous, and the second and third terms are weakly l.s.c., since they are norm l.s.c. and the space is convex. For the first term, let $u_n \xrightarrow{w} u$ in L_2. Then

$$\begin{aligned}
\|\nabla u\|_{L_2} &= \sup_{\psi\in(C_0^\infty)^d,\ \|\psi\|_{L^2}=1} |(\psi, \nabla u)_{L_2}| \\
&= \sup_{\psi\in(C_0^\infty)^d,\ \|\psi\|_{L^2}=1} |(\div\psi, u)_{L_2}| \\
&\le \sup_{\psi\in(C_0^\infty)^d,\ \|\psi\|_{L^2}=1} \lim_{n\to\infty} |(\div\psi, u_n)_{L_2}| \\
&= \sup_{\psi\in(C_0^\infty)^d,\ \|\psi\|_{L^2}=1} \lim_{n\to\infty} |(\psi, \nabla u_n)_{L_2}| \\
&\le \liminf_{n\to\infty} \|\nabla u_n\|_{L_2}
\end{aligned}$$

by Cauchy-Schwartz. Thus the first term is l.s.c. as well.

EXAMPLE (Geodesics). Let $M \subset \mathbb{R}^d$ be closed and let $\gamma : [0,1] \to M$ be a *rectifiable* curve (i.e., γ is continuous and γ', as a distribution, is in $L^1([0,1];\mathbb{R}^d)$). The *length* of γ is

$$L(\gamma) = \int_0^1 |\gamma'(s)|\, ds\ .$$

THEOREM 8.44. *Suppose $M \subset \mathbb{R}^d$ be closed. If $x, y \in M$ and there is at least one rectifiable curve $\gamma : [0,1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$, then there exists a rectifiable curve $\tilde\gamma : [0,1] \to M$ such that $\tilde\gamma(0) = x$, $\tilde\gamma(1) = y$, and*

$$L(\tilde\gamma) = \inf\{L(\gamma) | \gamma : [0,1] \to M \text{ is rectifiable and } \gamma(0) = x, \gamma(1) = y\}\ .$$

Such a minimizing curve is called a geodesic*.*

Note that a geodesic is the shortest path on some manifold M (i.e., surface in $\mathbb{R}^d$) between two points. One exists provided only that the two points can be joined within M. Note that a geodesic may not be unique (e.g., consider joining points $(-1,0)$ and $(1,0)$ within the unit circle).

PROOF. We would like to use Theorem 8.39; however, L^1 is *not* reflexive. We need two key ideas to resolve this difficulty. We expect that γ' is constant along a geodesic, so define

$$E(\gamma) = \int_0^1 |\gamma'(s)|^2\, ds$$

and let us try to minimize E in $L^2([0,1])$. This is the first key idea.

Define

$$Y = \left\{ f \in L^2([0,1];\mathbb{R}^d) : \gamma_f(s) \equiv x + \int_0^s f(t)\, dt \in M \text{ for all } s \in [0,1] \text{ and } \gamma_f(1) = y \right\}\ .$$

These are the derivatives of rectifiable curves from x to y. Since the map $f \mapsto \int_0^s f(t)\, dt$ is a continuous linear functional, Y is weakly closed in $L^2([0,1];\mathbb{R}^d)$. Since $\gamma_f' = f$, define $\tilde E : Y \to [0,\infty)$ by

$$\tilde E(f) = E(\gamma_f) = \int_0^1 |f(s)|^2\, ds\ .$$

Clearly $|\cdot|$ is convex, so $\tilde E$ is weakly l.s.c. by Lemma 8.42. Let

$$A_\alpha = \{f \in Y : \tilde E(f) \le \alpha\}\ ,$$

so that by definition A_α is bounded for any α. If A_α is not empty for some α, then there is a minimizer f_0 of $\tilde{E}$, by Theorem 8.39.

Now we need the second key idea. Given any rectifiable γ, define its *geodesic reparameterization* γ^* by

$$T(s) = \frac{1}{L(\gamma)} \int_0^s |\gamma'(t)|\, dt \in [0,1] \quad \text{and} \quad \gamma^*(T(s)) = \gamma(s) \ ,$$

which is well defined since T is nondecreasing and $T(s)$ is constant only where γ is also constant. But

$$\gamma'(s) = \big(\gamma^*(T(s))\big)' = \gamma^{*\prime}(T(s))\, T'(s) = \gamma^{*\prime}(T(s))\, \frac{|\gamma'(s)|}{L(\gamma)} \ ,$$

so

$$|\gamma^{*\prime}(s)| = L(\gamma)$$

is constant. Moreover, $L(\gamma^*) = L(\gamma)$, and so

$$E(\gamma^*) = L(\gamma^*)^2 \ .$$

Now at least one γ exists by hypothesis, so the reparameterized γ^* has $E(\gamma^*) < \infty$. Thus, for some α, A_α is nonempty, and we conclude that we have a minimizer f_0 of $\tilde{E}$.

Finally, for any rectifiable curve,

$$E(\gamma) \geq L(\gamma)^2 = L(\gamma^*)^2 = E(\gamma^*) \ .$$

Thus a curve of minimal energy E must have $|\gamma'|$ constant. So, for any rectifiable $\gamma = \gamma_f$ (where $f = \gamma'$),

$$L(\gamma) = E(\gamma_*)^{1/2} = \tilde{E}(f)^{1/2} \geq \tilde{E}(f_0)^{1/2} = E(\gamma_{f_0})^{1/2} = L(\gamma_{f_0}) \ ,$$

and γ_{f_0} is our geodesic. □

8.8. Exercises

1. Let X, Y_1, Y_2, and Z be normed linear spaces and $P : Y_1 \times Y_2 \to Z$ be a continuous bilinear map (so P is a "product" between Y_1 and Y_2).
   (a) Show that for $y_i, \hat{y}_i \in Y_i$,
   $$DP(y_1, y_2)(\hat{y}_1, \hat{y}_2) = P(y_1, \hat{y}_2) + P(\hat{y}_1, y_2) \ .$$
   (b) If $f : X \to Y_1 \times Y_2$ is differentiable, show that for $h \in X$,
   $$D(P \circ f)(x)\, h = P(Df_1(x)\, h, f_2(x)) + P(f_1(x), Df_2(x)\, h) \ .$$
2. Let X be a real Hilbert space and $A_1, A_2 \in B(X,X)$, and define $f(x) = (x, A_1 x)_X A_2 x$. Show that $Df(x)$ exists for all $x \in X$ by finding an explicit expression for it.
3. Let $X = C([0,1])$ be the space of bounded continuous functions on $[0,1]$ and, for $u \in X$, define $F(u)(x) = \displaystyle\int_0^1 K(x,y)\, f(u(y))\, dy$, where $K : [0,1] \times [0,1] \to \mathbb{R}$ is continuous and f is a C^1-mapping of $\mathbb{R}$ into $\mathbb{R}$. Find the Fréchet derivative $DF(u)$ of F at $u \in X$. Is the map $u \mapsto DF(u)$ continuous?
4. Suppose X and Y are Banach spaces, and $f : X \to Y$ is differentiable with derivative $Df(x) \in B(X,Y)$ being a compact operator for any $x \in X$. Prove that f is also compact.

5. Set up and apply the contraction mapping principle to show that the problem
$$-u_{xx} + u - \epsilon u^2 = f(x) , \quad x \in \mathbb{R} ,$$
has a unique smooth bounded solution if $\epsilon > 0$ is small enough, where $f(x) \in \mathcal{S}(\mathbb{R})$ is smooth and dies at infinity.

6. Use the contraction-mapping theorem to show that the Fredholm Integral Equation
$$f(x) = \varphi(x) + \lambda \int_a^b K(x,y) f(y)\, dy$$
has a unique solution $f \in C([a,b])$, provided that λ is sufficiently small, wherein $\varphi \in C([a,b])$ and $K \in C([a,b] \times [a,b])$.

7. Suppose that F is a compact linear operator on a Banach space X, that $x_0 = F(x_0)$ is a fixed point of F and that 1 is *not* an eigenvalue of $DF(x_0)$. Prove that x_0 is an isolated fixed point.

8. Consider the first-order differential equation
$$u'(t) + u(t) = \cos(u(t))$$
posed as an initial-value problem for $t > 0$ with initial condition
$$u(0) = u_0 .$$
(a) Use the contraction-mapping theorem to show that there is exactly one solution u corresponding to any given $u_0 \in \mathbb{R}$.

(b) Prove that there is a number ξ such that $\lim_{t\to\infty} u(t) = \xi$ for any solution u, independent of the value of u_0.

9. Set up and apply the contraction mapping principle to show that the boundary value problem
$$\begin{aligned} &-u_{xx} + u - \epsilon u^2 = f(x) , \quad x \in (0, +\infty) , \\ &u(0) = u(+\infty) = 0 , \end{aligned}$$
has a unique smooth solution if $\epsilon > 0$ is small enough, where $f(x)$ is a smooth compactly supported function on $(0, +\infty)$.

10. Consider the partial differential equation
$$\begin{aligned} &\frac{\partial u}{\partial t} - \frac{\partial^3 u}{\partial t \partial x^2} - \epsilon u^3 = f , \quad -\infty < x < \infty,\ t > 0 , \\ &u(x,0) = g(x) . \end{aligned}$$
Use the Fourier transform and a contraction mapping argument to show that there exists a solution for small enough ϵ. In what spaces should f and g lie?

11. Surjective Mapping Theorem: Let X and Y be Banach spaces, $U \subset X$ be open, $f : U \to Y$ be C^1, and $x_0 \in U$. If $Df(x_0)$ has a bounded right inverse, then $f(U)$ contains a neighborhood of $f(x_0)$.

(a) Prove this theorem from the Inverse Function Theorem. Hint: Let R be the right inverse of $Df(x_0)$ and consider $g : V \to Y$ where $g(y) = f(x_0 + Ry)$ and $V = \{y \in Y : x_0 + Ry \in U\}$.

(b) Prove that if $y \in Y$ is sufficiently close to $f(x_0)$, there is at least one solution to $f(x) = y$.

12. Let X and Y be Banach spaces.

(a) Let F and G take X to Y be C^1 on X, and let $H(x,\epsilon) = F(x) + \epsilon G(x)$ for $\epsilon \in \mathbb{R}$. If $H(x_0, 0) = 0$ and $DF(x_0)$ is invertible, show that there exists $x \in X$ such that $H(x,\epsilon) = 0$ for ϵ sufficiently close to 0.

(b) For small ϵ, prove that there is a solution $w \in H^2(0,\pi)$ to

$$w'' = w + \epsilon w^2 \ , \quad w(0) = w(\pi) = 0 \ .$$

13. Prove that for sufficiently small $\epsilon > 0$, there is at least one solution to the functional equation

$$f(x) + \sin x \int_{-\infty}^{\infty} f(x-y)\, f(y)\, dy = \epsilon e^{-|x|^2}, \quad x \in \mathbb{R} \ ,$$

such that $f \in L^1(\mathbb{R})$.

14. Let X and Y be Banach spaces, and let $U \subset X$ be open and convex. Let $F : U \to Y$ be an n-times Fréchet differentiable operator. Let $x \in U$ and $h \in X$. Prove that in Taylor's formula, the remainder is actually bounded as

$$\begin{aligned} \|R_{n-1}(x,h)\| &= \left\| F(x+h) - F(x) - DF(x)h + \cdots + \frac{1}{(n-1)!} D^{n-1}F(x)(h,\dots,h) \right\| \\ &\le \sup_{0\le\alpha\le1} \|D^n F(x+\alpha h)\| \, \|h\|^n \ . \end{aligned}$$

15. Prove that if X is a NLS, U a convex subset of X, and $f : U \to \mathbb{R}$ is strictly convex and differentiable, then, for $x, y \in U$, $x \neq y$,

$$f(y) > f(x) + Df(x)(y-x) \ ,$$

and $Df(x) = 0$ implies that f has a strict and therefore unique minimum.

16. Let $\Omega \subset \mathbb{R}^d$ have a smooth boundary, and let $g(x)$ be real with $g \in H^1(\Omega)$. Consider the BVP

$$\begin{cases} -\Delta u + u = 0, & \text{in } \Omega \ , \\ u = g, & \text{on } \partial\Omega \ . \end{cases}$$

(a) Write this as a variational problem.

(b) Define an appropriate energy functional $J(v)$ and find $DJ(v)$.

(c) Relate the BVP to a constrained minimization of $J(v)$.

17. Let $\Omega \subset \mathbb{R}^n$ have a smooth boundary, $A(x)$ be an $n \times n$ real matrix with components in $L^\infty(\Omega)$, and let $c(x)$, $f(x)$ be real with $c \in L^\infty(\Omega)$ and $f \in L^2(\Omega)$. Consider the BVP

$$\begin{cases} -\nabla\cdot A\nabla u + cu = f, & \text{in } \Omega \ , \\ u = 0, & \text{on } \partial\Omega \ . \end{cases}$$

(a) Write this as a variational problem.

(b) Assume that A is symmetric and uniformly positive definite and c is uniformly positive. Define the energy functional $J : H_0^1 \to \mathbb{R}$ by $J(v) = \dfrac{1}{2}\displaystyle\int_\Omega \big\{|A^{1/2}\nabla v|^2 + c|v|^2 - 2fv\big\}dx$. Find $DJ(v)$.

(c) Prove that for $u \in H_0^1$, the following are equivalent: (i) u is the solution of the BVP; (ii) $DJ(u) = 0$; (iii) u minimizes $J(v)$.

18. Let X and Y be Banach spaces, $U \subset X$ an open set, and $f : U \to Y$ Fréchet differentiable. Suppose that f is compact, in the sense that for any $x \in U$, if $\overline{B_r(x)} \subset U$, then $f(B_r(x))$ is precompact in Y. If $x_0 \in U$, prove that $Df(x_0)$ is a compact linear operator.

19. Let $F(u) = \displaystyle\int_{-1}^{5} [(u'(x))^2 - 1]^2 \, dx$.

    (a) Find all extremals in $C^1([-1,5])$ such that $u(-1) = 1$ and $u(5) = 5$.

    (b) Decide if any extremal from (a) is a minimum of F. Consider $u(x) = |x|$.

20. Consider the functional
$$F(y) = \int_0^1 [(y(x))^2 - y(x)\, y'(x)] \, dx,$$
defined for $y \in C^1([0,1])$.

    (a) Find all extremals.

    (b) If we require $y(0) = 0$, show by example that there is no minimum.

    (c) If we require $y(0) = y(1) = 0$, show that the extremal is a minimum. Hint: note that $y\, y' = (\frac{1}{2} y^2)'$.

21. Find all extremals of
$$\int_0^{\pi/2} \left[\big(y'(t)\big)^2 + \big(y(t)\big)^2 + 2\, y(t) \right] dt$$
under the condition $y(0) = y(\pi/2) = 0$.

22. Suppose that we wish to minimize
$$F(y) = \int_0^1 f(x, y(x), y'(x), y''(x)) \, dx$$
over the set of $y(x) \in C^2([0,1])$ such that $y(0) = \alpha$, $y'(0) = \beta$, $y(1) = \gamma$, and $y'(1) = \delta$. That is, with $C_0^2([0,1]) = \{u \in C^2([0,1]) : u(0) = u'(0) = u(1) = u'(1) = 0\}$, $y \in C_0^2([0,1]) + p(x)$, where p is the cubic polynomial that matches the boundary conditions.

    (a) Find a differential equation, similar to the Euler-Lagrange equation, that must be satisfied by the minimum (if it exists).

    (b) Apply your equation to find the extremal(s) of
$$F(y) = \int_0^1 (y''(x))^2 \, dx \ ,$$
where $y(0) = y'(0) = y'(1) = 0$ but $y(1) = 1$, and justify that each extremal is a (possibly nonstrict) minimum.

23. Prove the theorem: If f and g map $\mathbb{R}^3$ to $\mathbb{R}$ and have continuous partial derivatives up to second order, and if $u \in C^2([a,b])$, $u(a) = \alpha$ and $u(b) = \beta$, minimizes
$$\int_a^b f(x, u(x), u'(x)) \, dx \ ,$$
subject to the constraint
$$\int_a^b g(x, u(x), u'(x)) \, dx = 0 \ ,$$

then there is a nontrivial linear combination $h = \mu f + \lambda g$ such that $u(x)$ satisfies the Euler-Lagrange equation for h.

24. Consider the functional

$$\Phi(x, y, y') = \int_a^b F(x, y(x), y'(x))\, dx\ .$$

(a) If $F = F(y, y')$ only, prove that the Euler-Lagrange equations reduce to

$$\frac{d}{dx}(F - y' F_{y'}) = 0\ .$$

(b) Among all continuous curves $y(x)$ joining the points $(0, 1)$ and $(1, \cosh(1))$, find the one which generates the minimum area when rotated about the x-axis. Recall that this area is

$$A = 2\pi \int_0^1 y\sqrt{1 + (y')^2}\, dx\ .$$

$\left[\text{Hint: } \displaystyle\int \frac{dt}{\sqrt{t^2 - C^2}} = \ln(t + \sqrt{t^2 - C^2}).\right]$

25. Consider the functional

$$J[x, y] = \int_0^{\pi/2} [(x'(t))^2 + (y'(t))^2 + 2x(t)y(t)]\, dt$$

and the boundary conditions

$$x(0) = y(0) = 0 \quad \text{and} \quad x(\pi/2) = y(\pi/2) = 1\ .$$

(a) Find the Euler-Lagrange equations for the functional.

(b) Find all extremals.

(c) Find a global minimum, if it exists, or show it does not exist.

(d) Find a global maximum, if it exists, or show it does not exist.

26. Consider the problem of finding a C^1 curve that minimizes

$$\int_0^1 \left[(y'(t))^2 - 1\right] dt$$

subject to the conditions that $y(0) = y(1) = 0$ and

$$\int_0^1 (y(t))^2\, dt = 1\ .$$

(a) Remove the integral constraint by incorporating a Lagrange multiplier, and find the Euler equations.

(b) Find all extremals to this problem.

(c) Find the solution to the problem.

(d) Use your result to find the best constant C in the inequality

$$\|y\|_{L^2(0,1)} \le C\|y'\|_{L^2(0,1)}$$

for functions that satisfy $y(0) = y(1) = 0$.

27. Find the C^2 curve $y(t)$ that minimizes the functional

$$\int_0^1 \left[(y(t))^2 + (y'(t))^2\right] dt$$

subject to the endpoint constraints

$$y(0) = 0 \quad \text{and} \quad y(1) = 1$$

and the constraint

$$\int_0^1 y(t)\, dt = 0 \ .$$

28. Find the form of the curve in the plane (*not* the curve itself), of minimal length, joining $(0,0)$ to $(1,0)$ such that the area bounded by the curve, the x and y axes, and the line $x = 1$ has area $\pi/8$.

29. Solve the constrained Brachistochrone problem: In a vertical plane, find a C^1-curve joining $(0,0)$ to (b,β), b and β positive and given, such that if the curve represents a track along which a particle slides without friction under the influence of a constant gravitational force of magnitude g, the time of travel is minimal. Note that this travel time is given by the functional

$$T(y) = \int_0^b \sqrt{\frac{1 + (y'(x))^2}{2g(\beta - y(x))}}\, dx \ .$$

30. Consider a stream between the lines $x = 0$ and $x = 1$, with speed $v(x)$ in the y-direction. A boat leaves the shore at $(0,0)$ and travels with constant speed $c > 0$. The problem is to find the path $y(x)$ of minimal crossing time, where the terminal point $(1,\beta)$ is unspecified.

(a) Find conditions on y so that it satisfies the Euler-Lagrange constraint. Hint: the crossing time is

$$t = \int_0^1 \frac{\sqrt{c^2(1 + (y')^2) - v^2} - vy'}{c^2 - v^2}\, dx \ .$$

(b) What free endpoint constraint (transversality condition) is required?

(c) If v is constant, find y.

Bibliography

[Ad] R.A. Adams, *Sobolev Spaces*, Academic Press, 1975.

[Au] J.-P. Aubin, *Applied Functional Analysis*, Wiley, 1979.

[Car] C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the first Order*, 1982.

[Ch] E.W. Cheney, *Notes on Applied Mathematics*, Department of Mathematics, The University of Texas at Austin, Austin, TX, 1998.

[DM] L. Debnath & P. Mikusiński, *Introduction to Hilbert Spaces with Applications*, Academic Press, 1990.

[Fo] G.B. Folland, *Introduction to Partial Differential Equations*, Princeton, 1976.

[GF] I.M. Gelfand & S.V. Fomin, *Calculus of Variations*, Prentice Hall, 1963.

[GT] D. Gilbarg & N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Spring-Verlag, 1983.

[Ho] M.H. Holmes, *Introduction to Perturbation Methods*, Springer-Verlag, New York, 1995.

[JLJ] J. Jost & X. Li-Jost, *Calculus of Variations*, Cambridge, 1998.

[Ko] H.A. Koch, *Notes on Applied Mathematics*, Department of Mathematics, The University of Texas at Austin, Austin, TX, 1998.

[Kr] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1978.

[LL] E.H. Lieb & M. Loss, *Analysis*, AMS, 1977.

[MZ] D. Mitrović and D. Žubrinić, *Fundamentals of Applied Functional Analysis*, Addison Wesley Longman, 1998.

[OD] J.T. Oden & L.F. Demkowicz, *Applied Functional Analysis*, CRC Press, 1996.

[Ol] F.W.J. Olver, *Asymptotics and Special Functions*, Academic Press, 1974.

[RS] M. Reed & B. Simon, *Methods of Modern Physics, Vol.* 1*, Functional analysis*, Academic Press, 1980.

[Roy] H.L. Royden, *Real Analysis*, 3rd ed., MacMillan Publishing Co., New York, 1988.

[Ru0] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 1976.

[Ru1] W. Rudin, *Functional Analysis*, McGraw-Hill, New York, 1991.

[Ru2] W. Rudin, *Real and Complex Analysis*, 3*rd* Ed., McGraw-Hill, New York, 1987.

[Sa] H. Sagan, *Introduction to the Calculus of Variations*, Dover, 1969.

[Sho] R.E. Showalter, *Hilbert Space Methods for Partial Differential Equations*, 1994 (available on the internet at `http://ejde.math.swt.edu/mono-toc.html`).

[Si] A.H. Siddiqi, *Applied Functional Analysis*, Marcel Dekker, New York, 2004.

[St] E. Stein, *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, 1970.

[SW] E. Stein & G. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, 1971.

[Yo] K. Yosida, *Functional Analysis*, Springer-Verlag, New York, 1980.