MATH 5410: INTRODUCTION TO APPLIED MATHEMATICS

ALEC WENDLAND

1. Normed Linear Spaces

This section gives an introduction to normed linear spaces. By a *linear space* we mean that any linear combination of functions in the family will be another member of the family. By a *normed* linear space we mean any linear space endowed with a norm.

1.1. **Definitions and Examples.**

Definition 1.1.1 (Real Vector Space (Real Linear Space)). *A* **real vector space** *(or* **real linear space***) is a triple $(X, +, \cdot)$ in which X is a set, and $+$ and $\cdot$ are binary operations satisfying the following ten axioms. Here we assume that $x, y, z \in X$ and $\lambda, \mu \in \mathbb{R}$.*

(1) $x + y \in X$ *(closure)*
(2) $x + y = y + x$ *(commutativity)*
(3) $x + (y + z) = (x + y) + z$ *(associativity)*
(4) There exists a unique element $0 \in X$ *such that* $x + 0 = x$ *for all* $x \in X$ *(additive identity)*
(5) For each $x \in X$ *there is a unique element* $(-x) \in X$ *such that* $x + (-x) = 0$ *(additive inverse)*
(6) $\lambda \cdot x \in X$ *(closure)*
(7) $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$ *(distributivity)*
(8) $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$ *(distributivity)*
(9) $\lambda \cdot (\mu \cdot x) = (\lambda \cdot \mu) \cdot x$ *(associativity)*
(10) $1 \cdot x = x$ *(multiplicative identity/unity).*

Note that the first five axioms define an additive abelian group. In axiom (4), note that the uniqueness of 0 need not be mentioned, as it may be proved from axiom (2). To see this, suppose that there are two additive identities, say 0^* and 0^{**}. Then

$$0^* = 0^* + 0^{**} = 0^{**} + 0^* = 0^{**}.$$

Also note that a consequence of axiom (7) is

$$\lambda \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \lambda x_i.$$

We can also define a *complex vector space* X. In such a space λx is defined and is an element of X whenever $\lambda \in \mathbb{C}$ and $x \in X$. We call the field elements λ the *scalars* and the elements x of X the *vectors*.

Definition 1.1.2 (Norm). *Let X be a linear space. A* **norm** *on X is a real–valued function, denoted by $\| \cdot \|$, that satisfies the following three axioms:*

(1) $\|x\| > 0$ *for all* $x \in X$ *such that* $x \neq 0$;

Date: August 14, 2022.

(2) $\|\lambda x\| = |\lambda|\|x\|$ *for all* $\lambda \in \mathbb{R}$ *and* $x \in X$.
(3) $\|x + y\| \leq \|x\| + \|y\|$ *for all* $x, y \in X$ *(triangle inequality).*

Definition 1.1.3 (Normed Linear Space)**.** *A linear space* $(X, +, \cdot)$ *on which a norm has been introduced is called a* **normed linear space***.*

Example 1.1.4. *Let* $X = \mathbb{R}$*, and define* $\|x\| = |x|$*.*

Example 1.1.5. *Let* $X = \mathbb{C}$*, with scalar field also* $\mathbb{C}$*. Define* $\|x\| = |x|$*, where* $|x|$ *denotes the modulus of* x*, that is, if* $x = a + ib$*,* $|x| = \sqrt{a^2 + b^2}$*.*

Example 1.1.6. *Let* $X = \mathbb{C}$ *and take the scalar field to be* $\mathbb{R}$*. Note that this is now a* **real** *vector space.*

Example 1.1.7. *Let* $X = \mathbb{R}^n$*. Here the elements of* X *are* $n-$*tuples of real numbers*

$$x = [x_1, x_2, \ldots, x_n]^\top.$$

One possible norm is defined by

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|.$$

This is called the supremum/uniform/infinity norm.

Example 1.1.8. *Let* $X = \mathbb{R}^n$*, and define a norm by*

$$\|x\| = \sum_{i=1}^{n} |x_i|.$$

Note that in this example and the previous example we have two distinct normed linear spaces, but each involves the same linear space. Thus we may refer to a normed linear space as a pair $(X, \|\cdot\|)$*, in which we specify the norm under consideration.*

Example 1.1.9. *Let* $X = \mathcal{C}[a, b]$*, the set of all real–valued continuous functions defined on a fixed compact interval* $[a, b]$*. The supremum/infinity/uniform/maximum norm is defined by*

$$\|x\|_\infty = \max_{a \leq t \leq b} |x(t)|.$$

Example 1.1.10. *Let* X *be the set of all Lebesgue–integrable functions defined on a fixed compact interval* $[a, b]$*. The usual norm for this space is*

$$\|x\| = \int_a^b |x(t)| \, dt.$$

Note that in this space, the vectors are actually equivalence classes of functions, with two functions being regarded as equivalent if they differ only on a set of measure zero.

Example 1.1.11. *Let* $X = \ell$*, the space of all sequences*

$$\{x_n\}_{n=1}^{\infty} = \{x_1, x_2, \ldots, \}$$

in which only a finite number of terms are nonzero. Define

$$\|x\| = \max_n |x_n|.$$

Example 1.1.12. *Let* $X = \ell_\infty$*, the space of all real sequences* x *for which* $\sup_n |x_n| < \infty$*. Define*

$$\|x\| = \sup_n |x_n|.$$

Example 1.1.13. *Let $X = \Pi$, the space of all polynomials having real coefficients. A typical element of Π is a function x having the form*

$$x(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n.$$

One possible norm on Π is $\|x\| = \max_i |a_i|$. Others are $\|x\| = \max_{0 \le t \le 1} |x(t)|$ or $\|x\| = \int_0^1 |x(t)|\, dt$ or $\|x\|_p = \left(\sum\limits_{i=0}^{n} |x|^p\right)^{1/p}$.

Example 1.1.14. *Let $X = \mathbb{R}^n$ and use the* **Euclidean norm***, defined by*

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}.$$

We recall the concept of linear independence.

Definition 1.1.15 (Linearly Independent)**.** *A subset S in a linear space is* **linearly independent** *if it is not possible to find a finite, nonempty set of distinct vectors $x_1, x_2, \ldots, x_m \in S$ and nonzero scalars $c_1, c_2, \ldots, c_m$ such that*

$$c_1 x_1 + c_2 x_2 + \cdots + c_m x_m = 0.$$

Definition 1.1.16 (Linearly Dependent)**.** *A subset S in a linear space is* **linearly dependent** *if it is not linearly independent.*

We recall the following from linear algebra.

Definition 1.1.17 (Span)**.** *The* **span** *of a set S in a (real) linear space X is the set of all vectors in X that are expressible as linear combinations of vectors in S. More precisely,*

$$\operatorname{span}(S) := \left\{\sum_{i=1}^{m} \lambda_i x_i : x_i \in S, \lambda_i \in \mathbb{R}, i = 1, 2, \ldots, m\right\}.$$

Recall that linear combinations are always finite expressions of the form

$$\sum_{i=1}^{n} \lambda_i x_i.$$

We say that S *spans* X when $\operatorname{span}(S) = X$.

Definition 1.1.18 (Basis)**.** *A* **basis** *for a linear space X is any set that is linearly independent and spans X.*

Note that both properties here are needed. Any set that is linearly independent is contained in a basis, and any set that spans the space contains a basis.

Definition 1.1.19 (Finite Dimensional)**.** *A linear space is said to be* **finite dimensional** *if it has a finite basis.*

If a linear space is finite dimensional, then every basis for that space has the same number of elements.

Definition 1.1.20 (Dimension)**.** *The number of elements in any basis for a finite dimensional linear space X is called the* **dimension** *of X, denoted by* $\dim(X)$.

1.2. **Convexity, Convergence, Compactness, and Completeness.**

Definition 1.2.1 (Convex)**.** *A subset K in a linear space X is said to be* **convex** *if it contains every straight line segment connecting any two of its elements. More precisely, K is convex if for any $x, y \in K$ and $\lambda \in [0,1]$, we have $\lambda x + (1-\lambda)y \in K$.*

The notion of convexity arises frequently in optimization problems. For instance, much of linear programming is based on the fact that a linear function on a convex polyhedral set must attain its extreme values at the vertices of the set.

Example 1.2.2 (Examples of convex sets)**.** *Let X be a linear space. Then the following sets are convex in X :*

(1) The space X itself;
(2) Any single point set (trivially);
(3) the empty set $\emptyset$;
(4) Any linear subspace of X;
(5) Any straight line segment, that is, a set of the following form for a fixed $a, b \in X$:

$$\{\lambda a + (1-\lambda)b : 0 \leq \lambda \leq 1\}.$$

Definition 1.2.3 (Unit Ball)**.** *Let $(X, \|\cdot\|)$ be a normed linear space. Then the* **unit ball** *is defined by*

$$B[0,1] := \{x \in X : \|x\| \leq 1\}.$$

Proposition 1.2.4 (Unit Ball is Convex)**.** *In any normed linear space $(X, \|\cdot\|)$, the unit ball $B[0,1]$ is convex.*

Proof. Let $x, y \in B[0,1]$ and note then that $\|x\| \leq 1$ and $\|y\| \leq 1$. Let $\lambda \in [0,1]$. Observe

$$\begin{aligned} \|\lambda x + (1-\lambda)y\| &\leq \|\lambda x\| + \|(1-\lambda)y\| \\ &= |\lambda|\|x\| + |1-\lambda|\|y\| \\ &= \lambda\|x\| + (1-\lambda)\|y\| \\ &\leq \lambda + (1-\lambda) \\ &= 1. \end{aligned}$$

□

Definition 1.2.5 (Metric)**.** *let X be a set. A* **metric** *$d : X \times X \to \mathbb{R}$ is a real–valued function on X such that for all $x, y, z \in X$,*

(1) $d(x,y) \geq 0$;
(2) $d(x,y) = 0$ if and only if $x = y$;
(3) $d(x,y) = d(y,x)$;
(4) $d(x,y) \leq d(x,z) + d(z,y)$.

Definition 1.2.6 (Metric Space)**.** *A* **metric space** *is a pair (X, d) in which X is a set and d is a metric on X.*

Proposition 1.2.7 (Metric Space Induced By Norm)**.** *Let $(X, \|\cdot\|)$ be a normed linear space. Then the function $d : X \times X \to \mathbb{R}$ defined by*

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

defines a metric on X. Moreover, (X, d) is a metric space.

Proof. Properties (1), (2), and (3) are immediate. To see (4), let $x, y, z \in X$. Then

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

□

Definition 1.2.8 (Convergent Sequence). *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a normed linear space X. We say that $\{x_n\}_{n\in\mathbb{N}}$* **converges** *to a point $x \in X$ (and write $x_n \to x$) if*

$$\lim_{n\to\infty} \|x_n - x\| = 0.$$

Example 1.2.9. *Let $X = (\mathcal{C}[0,1], \|\cdot\|_\infty)$ and consider the sequence of functions defined by*

$$x_n(t) := \sin(t/n).$$

Then $\{x_n\}_{n\in\mathbb{N}}$ converges to zero, for observe that

$$\|x_n - 0\|_\infty = \sup_{0\leq t\leq 1} |\sin(t/n)| = \sin(1/n) \to 0$$

as $n \to \infty$.

The concept of convergence is often needed in applied mathematics. For instance, the solution to a particular problem may be difficult (or impossible) to obtain but may be approached by a suitable sequence of functions that are easier to obtain and may possibly be calculated explicitly.

Definition 1.2.10 (Subsequence). *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a space X. A* **subsequence** *of $\{x_n\}_{n\in\mathbb{N}}$ is a sequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ where the integers n_k, $k \in \mathbb{N}$ are such that $n_1 < n_2 < n_3 < \dots$ and*

$$\{x_{n_k} : k \in \mathbb{N}\} \subseteq \{x_n : n \in \mathbb{N}\}.$$

Definition 1.2.11 (Compactness). *A subset K in a normed linear space X is said to be* **compact** *if every sequence $x_{n n\in\mathbb{N}}$ in K has a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ that converges to a point in K.*

Note that this definition coincides with the definition of *sequential compactness* from general topology.

We recall the following definitions from real analysis.

Definition 1.2.12 (Bounded Below (Above)). *Let T be a set of real numbers. We say that T is* **bounded below (above)** *if there exists $M \in \mathbb{R}$ such that $t \geq M$ ($t \leq M$) for all $t \in T$.*

Definition 1.2.13 (Infimum, Greatest Lower Bound). *Let T be a set of real numbers. A number $\tilde{b}$ is called the* **infimum** *or* **greatest lower bound** *of T if $\tilde{b}$ satisfies the following two properties:*

(1) $\tilde{b}$ is a lower bound for T;
(2) If b is a lower bound for T, then $\tilde{b} \geq b$.

We write $\tilde{b} =: \inf T$.

Definition 1.2.14 (Supremum, Least Upper Bound). *Let T be a set of real numbers. A number $\tilde{b}$ is called the* **supremum** *or* **least upper bound** *of T if $\tilde{b}$ satisfies the following two properties:*

(1) $\tilde{b}$ is an upper bound for T;
(2) If b is an upper bound for T, then $\tilde{b} \leq b$.

We write $\widetilde{b} =: \sup T$.

Recall that the completeness axiom of $\mathbb{R}$ states that if T is nonempty and bounded below (above), then $\inf T$ ($\sup T$) always exists.

Definition 1.2.15 (Distance From a Set)**.** *Let X and Y be normed linear spaces and let $x \in X$. The* **distance** *from x to Y is defined to be the number*

$$dist(x, Y) := \inf_{y \in Y} \|x - y\|.$$

Theorem 1.2.16. *Let K be a compact subset of a normed linear space X. To each point $x \in X$ there corresponds at least one point $z^* \in K$ of minimum distance from x. More precisely, for all $x \in X$, there exists $z^* \in K$ such that*

$$\|z^* - x\| \leq \|z - x\|$$

for all $z \in K$.

Proof. Let $x \in X$ be any member of X. By definition of an infimum, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in K converging to $\operatorname{dist}(x, K)$. Since K is compact, there exists a convergent subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ converging to a point in K, say $z \in K$. By the triangle inequality, we have

$$\|x - z\| \leq \|x - x_{n_k}\| + \|x_{n_k} - z\|,$$

and, letting $k \to \infty$, we have

$$\operatorname{dist}(x, K) \leq \|x - z\| \leq \operatorname{dist}(x, K).$$

That is,

$$\|x - z\| = \operatorname{dist}(x, K).$$

This completes the proof. □

Example 1.2.17. *In $\mathbb{R}$, the open interval (a, b) is not compact, for we can take a sequence in the interval that converges to the endpoint b, say. Then every subsequence also converges to b, but since b is not in the interval, the interval cannot be compact.*

On the other hand, the closed and bounded interval $[a, b]$ is compact by the Heine–Borel Theorem.

Consider a sequence $\{x_n\}_{n\in\mathbb{N}}$ in a general metric space (X, d) and the associated sequence $\{d(x_n, x_m)\}_{n,m\in\mathbb{N}}$. Note that for any $x \in X$, we may write

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m).$$

Thus if $\{d(x_n, x_m)\}_{n,m\in\mathbb{N}}$ does not converge to zero, then $\{x_n\}_{n\in\mathbb{N}}$ cannot converge. On the other hand, if $\{d(x_n, x_m)\}_{n,m\in\mathbb{N}}$ does converge to zero, we can only guarantee that the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges if X is *complete.*

Definition 1.2.18 (Cauchy Sequence)**.** *Let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in a normed linear space X. We say that $\{x_n\}_{n\in\mathbb{N}}$ is a* **Cauchy sequence** *if*

$$\lim_{N\to\infty} \sup_{\substack{n\geq N \\ m\geq N}} \|x_n - x_m\| = 0.$$

Definition 1.2.19 (Complete Normed Linear Space)**.** *Let X be a normed linear space. If every Cauchy sequence in X is convergent, then the space X is said to be* **complete**.

Definition 1.2.20 (Banach Space)**.** *A complete normed linear space is called a* **Banach space**.

Note that completeness is important in constructing solutions to a problem by taking the limit of successive approximations. We often want information about the limit of these approximations, that is, the solution. For instance, if all the approximating functions are continuous and bounded, we should expect the limit of these functions to also be continuous and bounded. These properties depend on the norm that has been chosen and the function space that goes with it. Typically, we want a norm that leads to a Banach space.

Example 1.2.21. *Define $X = (\ell, \|\cdot\|_\infty)$. Further, define a sequence $\{x_n\}_{n\in\mathbb{N}}$ in ℓ as follows:*

$$x_n = \left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, 0, 0, \ldots\right\}.$$

If $m > n$, then

$$x_m - x_n = \left\{0, 0, \ldots, 0, \frac{1}{n+1}, \frac{1}{n+2}, \ldots, \frac{1}{m}, 0, \ldots\right\}.$$

Since $\|x_m - x_n\|_\infty = 1/(n+1)$, we find that the sequence $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy.

To see that this space is not complete, suppose by contradiction that $x_n \to y$ for some $y \in \ell$. The point y, being an element of ℓ, would be finitely nonzero, say $y_n = 0$ for $n \geq N$ for some positive integer N. Then for $m > N$, x_m would have as its $N-th$ term the value $1/N$, which the $N-th$ term of y is zero. Thus $\|x_m - y\|_\infty \geq 1/N$, and thus we cannot have convergence.

Theorem 1.2.22 (Continuous Functions on Closed Intervals Form a Banach Space)**.** *The normed linear space $(\mathcal{C}[a,b], \|\cdot\|_\infty)$ is a Banach space.*

Proof. Recall that $(\mathcal{C}[a,b], \|\cdot\|_\infty)$ is a normed linear space. Thus, it suffices to show that $(\mathcal{C}[a,b], \|\cdot\|_\infty)$ is complete.

Let $\{x_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $(\mathcal{C}[a,b], \|\cdot\|_\infty)$. Then for any $t \in [a,b]$, $\{x_n(t)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. By the completeness of $\mathbb{R}$, evidently $\{x_n(t)\}_{n\in\mathbb{N}}$ converges to a real number, say $\lim_{n\to\infty} x_n(t) =: x(t)$. We show that x is a continuous function on $[a,b]$ such that $\|x_n - x\|_\infty \to 0$.

We first show that $\{x_n\}_{n\in\mathbb{N}}$ converges to x uniformly on $[a,b]$. Fix $\epsilon > 0$. Since $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy, there exists a positive integer N such that for all $n, m \geq N$, we have

$$|x_n(t) - x_m(t)| < \epsilon.$$

Fix n. Letting $m \to \infty$, we get $x_m(t) \to x(t)$, so that

$$\lim_{m\to\infty} |x_n(t) - x_m(t)| = |x_n(t) - x(t)| < \epsilon$$

for every $n \geq N$ and $t \in [a,b]$. Thus $\{x_n\}_{n\in\mathbb{N}}$ converges uniformly on $[a,b]$.

To show convergence in norm, fix $\epsilon > 0$. Since $x_n \to x$ uniformly on $[a,b]$, there exists a positive integer N such that for all $t \in [a,b]$ and $n \geq N$, we have

$$|x_n(t) - x(t)| < \epsilon/2.$$

In particular, since this is for all $t \in [a,b]$, it follows

$$\|x_n - x\|_\infty = \sup_{t\in[a,b]} |x_n(t) - x(t)| \leq \epsilon/2 < \epsilon.$$

This shows $\|x_n - x\|_\infty \to 0$.

Finally, we show that x is continuous. Fix $\epsilon > 0$ and $t \in [a, b]$. By the uniform convergence of $\{x_n\}_{n\in\mathbb{N}}$, there exists a positive integer N such that for any $z \in [a, b]$, we have

$$|x_n(z) - x(z)| < \epsilon/3$$

for all $n \geq N$. Moreover, since x_N is continuous, there exists $\delta > 0$ such that for all $s \in [a, b]$ satisfying $|s - t| < \delta$, we have

$$|x_N(s) - x_N(t)| < \epsilon/3.$$

Thus, for any $s \in [a, b]$ such that $|s - t| < \delta$, it follows

$$\begin{aligned}|x(s) - x(t)| &\leq |x(s) - x_N(s)| + |x_N(s) - x_N(t)| + |x_N(t) - x(t)| \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 \\ &= \epsilon.\end{aligned}$$

Since ϵ and t were arbitrary, it follows that x is continuous. This completes the proof. □

Note that the traditional formulation of the theorem (1.2.22) states that a uniformly convergent sequence of continuous functions on a compact set must have a continuous limiting function.

Definition 1.2.23 (Uniform Convergence)**.** *A sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ on a set X is said to* **converge uniformly** *to a function f if for all $\epsilon > 0$ there exists a positive integer $N(\epsilon)$ such that for all $n \geq N$ and $x \in X$, we have*

$$|f_n(x) - f(x)| < \epsilon.$$

Formally,

$$\forall\epsilon \quad \exists N \quad \forall n \quad \forall x \quad [n \geq N \implies |f_n(x) - f(x)| < \epsilon].$$

Note that we may also write

$$\forall\epsilon \quad \exists N \quad \forall n \quad [n \geq N \implies \|f_n - f\|_\infty < \epsilon].$$

Definition 1.2.24 (Pointwise Convergence)**.** *A sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ on a set X is said to* **converge pointwise** *to a function f if for all $x \in X$ and any $\epsilon > 0$, there exists a positive integer $N(x, \epsilon)$ such that for all $n \geq N$, we have*

$$|f_n(x) - f(x)| < \epsilon.$$

Formally,

$$\forall x \quad \forall\epsilon \quad \exists N \quad \forall n \quad [n \geq N \implies |f_n(x) - f(x)| < \epsilon].$$

Example 1.2.25. *Consider the sequence of functions $\{f_n\}_{n\in\mathbb{N}}$ defined by*

$$f_n(x) := \begin{cases} |nx - 1|, & [0, 2/n] \\ 1, & \textit{otherwise.} \end{cases}$$

Then $\{f_n\}_{n\in\mathbb{N}}$ converges pointwise but not uniformly.

1.3. **Continuity, Open Sets, Closed Sets.**

Definition 1.3.1 (Continuity at a Point). *A function $f : D \subseteq X \to Y$ from a normed linear space X into a normed linear space Y is said to be* **continuous** *at a point $x \in D$ if for every sequence $\{x_n\}_{n\in\mathbb{N}}$ in D converging to x, we have also that $\{f(x_n)\}_{n\in\mathbb{N}}$ converges to $f(x)$.*

Definition 1.3.2 (Continuity on a Set). *A function $f : D \subseteq X \to Y$ is said to be* **continuous** *throughout D if f is continuous at every point $x \in D$.*

We see that a continuous function is one that preserves the convergence of sequences.

Example 1.3.3. *In any normed linear space, the norm $\|\cdot\|$ is continuous. To see this, let $\{x_n\}_{n\in\mathbb{N}}$ converge to x in a normed linear space X. Fix $\epsilon > 0$. There exists $N \in \mathbb{N}$ such that for all $n \geq N$, we have*

$$\|x_n - x\| < \epsilon.$$

By the reverse triangle inequality, it follows

$$|\|x_n\| - \|x\|| \leq \|x_n - x\| < \epsilon.$$

This completes the proof.

Theorem 1.3.4 (Continuous Functions Preserve Compactness). *Let $f : D \subseteq X \to Y$ be a continuous function from a normed linear space X into a normed linear space Y. If D is compact, then $f(D)$ is compact.*

Proof. Let $\{y_n\}_{n\in\mathbb{N}}$ be any sequence in $f(D)$. By definition of $f(D)$, there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in D such that for all $n \in \mathbb{N}$, we have $f(x_n) = y_n$. Since D is compact by the assumption, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ converging to a point $x \in D$. Since f is continuous, it follows

$$f(x) = f\left(\lim_{k\to\infty} x_{n_k}\right) = \lim_{k\to\infty} f(x_{n_k}) = \lim_{k\to\infty} y_{n_k}.$$

Since $f(x) \in f(D)$, evidently the sequence $\{y_{n_k}\}_{k\in\mathbb{N}}$ converges to a point in $f(D)$. This completes the proof. □

The following is a generalization of the extreme value theorem.

Theorem 1.3.5 (Generalized Extreme Value Theorem). *Let $f : D \subseteq X \to \mathbb{R}$ be a continuous real–valued function from a normed linear space X into $\mathbb{R}$ with D a compact subset of X. Then f attains its supremum and infimum, that is, there exist points $\hat{x}_1, \hat{x}_2 \in D$ such that*

$$f(\hat{x}_1) = \sup_{x\in D} f(x), \quad f(\hat{x}_2) = \inf_{x\in D} f(x).$$

Proof. Let $f : D \subseteq X \to \mathbb{R}$ be a continuous function with D a compact subset of a normed linear space X. Put

$$M := \sup_{x\in D} f(x).$$

There exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in D such that $\{f(x_n)\}_{n\in\mathbb{N}}$ approaches M. Since D is compact, it follows that $f(D)$ is also compact (1.3.4), so that there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ converging to a point $\hat{x}_1 \in D$. Moreover, since f is continuous, it follows that

$$f(\hat{x}_1) = f\left(\lim_{k\to\infty} x_{n_k}\right) = \lim_{k\to\infty} f(x_{n_k}).$$

By uniqueness of limits, $f(x) = M$, and thus $M < \infty$.

The proof of the infimum is similar. □

Definition 1.3.6 (Uniform Continuity)**.** *A function $f : X \to Y$ from a normed linear space X into a normed linear space Y is said to be* **uniformly continuous** *if for every $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that for all $x, y \in X$ satisfying $\|x-y\|_X < \delta$, we have $\|f(x)-f(y)\|_Y < \epsilon$.*

The important aspect of this definition is that δ depends only on ϵ, but not on x or y.

Theorem 1.3.7 (Continuous Functions on Compact Sets are Uniformly Continuous)**.** *Let $f : D \subset X \to Y$ be a continuous function from a normed linear space X into normed linear space Y with D compact. Then f is uniformly continuous throughout D.*

Proof. Let $f : D \subset X \to Y$ be a continuous function from a normed linear space X into a normed linear space Y with D compact and, by contradiction, suppose that f is not uniformly continuous.

Since f is not uniformly continuous, there exists $\epsilon > 0$ for which there exists no such $\delta > 0$ such that $\|f(x) - f(y)\|_Y < \epsilon$ whenever $\|x - y\|_X < \delta$. Thus, for each $n \in \mathbb{N}$, there exists a pair (x_n, y_n) of points of D satisfying the condition $\|x_n - y_n\|_X < \delta$ and $\|f(x_n)-f(y_n)\|_Y \geq \epsilon$. Since D is compact, the sequence $\{x_n\}_{n\in\mathbb{N}}$ has a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ converging to a point $x \in D$. Then $y_{n_k} \to x$ also, for observe that

$$\|y_{n_k} - x\|_X \leq \|y_{n_k} - x_{n_k}\|_X + \|x_{n_k} - x\|_X.$$

But then, by the continuity of f, we see that for any $n_k \in \mathbb{N}$,

$$\begin{aligned}\epsilon &\leq \|f(x_{n_k}) - f(y_{n_k})\|_Y \\ &\leq \|f(x_{n_k}) - f(x)\|_Y + \|f(x) - f(y_{n_k})\|_Y.\end{aligned}$$

Thus, letting $n_k \to \infty$, we get that $\epsilon \leq 0$, a contradiction. □

Definition 1.3.8 (Closed Set)**.** *A subset F in a normed linear space is said to be* **closed** *if the limit of every convergent sequence in F is also in F. Formally,*

$$[x_n \in F \quad \& \quad x_n \to x] \implies x \in F.$$

Proposition 1.3.9. *The intersection of any family of closed sets in a normed linear space is closed.*

Proof. Let $\{F_\alpha\}_{\alpha\in A}$ be a collection of closed subsets of a normed linear space X. Put $F := \bigcap_{\alpha\in A} F_\alpha$ and let $\{x_n\}_{n\in\mathbb{N}}$ be a convergent sequence in F, say $x_n \to x$ as $n \to \infty$. Since F_α is closed for each $\alpha \in A$, we have by definition that $x \in F_\alpha$ for all α. Hence, $x \in F$ and we conclude that F is closed. □

Definition 1.3.10 (Closure of a Set)**.** *Let A be a subset of a normed linear space X and let $\{F_\gamma\}_{\gamma\in\Gamma}$ be the set of all closed subsets of X containing A as a subset. Then the* **closure** *of A, denoted $\bar{A}$, is defined by*

$$\bar{A} := \bigcap_{\gamma\in\Gamma} F_\gamma.$$

We see immediately that for any set A in a normed linear space, $\bar{A}$ is a closed set containing A and is the smallest such closed set containing A.

We recall the definition of a *preimage.*

Definition 1.3.11 (Preimage)**.** *Let $f : X \to Y$ be a map from a set X into a set Y and let $A \subseteq Y$. Then the* **preimage** *of A under f is defined by the set*

$$f^{-1}(A) := \{x \in X : f(x) \in A\}.$$

Theorem 1.3.12. *Let $f : X \to Y$ be a continuous function from a normed linear space X into a normed linear space Y. Then for any closed set $F \in Y$, the inverse image $f^{-1}(F)$ of F is closed in X.*

Proof. Let F be a closed subset of Y and let $\{x_n\}_{n\in\mathbb{N}}$ be a convergent sequence in $f^{-1}(F)$, say, $x_n \to x \in X$ as $n \to \infty$. Since f is continuous, it follows that $f(x_n) \to f(x)$ as $n \to \infty$, and, since F is closed, we have $f(x) \in F$. Hence $x \in f^{-1}(F)$ and we see that $f^{-1}(F)$ is closed. $\square$

Example 1.3.13. *Let $(X, \|\cdot\|)$ be a normed linear space and consider the unit ball*

$$B[0,1] := \{x \in X : \|x\| \leq 1\}.$$

Noting that $\|\cdot\|^{-1}[0,1] = B[0,1]$ and $\|\cdot\|$ has been shown to be continuous, it follows that the unit ball is closed in any normed linear space.

Likewise, the following sets are closed in any normed linear space:

(1) $\{x \in X : \|x - a\| \leq r\}$;
(2) $\{x \in X : \|x - a\| \geq r\}$;
(3) $\{x \in X : r \leq \|x - a\| \leq s\}$.

Definition 1.3.14 (Open Set)**.** *A subset A of a normed linear space X is said to be* **open** *if its complement A^C is closed in X.*

Example 1.3.15. *In any normed linear space $(X, \|\cdot\|)$, the* **open unit ball**

$$B(0,1) := \{x \in X : \|x\| < 1\}$$

is open.

Likewise, the following sets are open in any normed linear space:

(1) $\{x \in X : \|x - a\| > r\}$;
(2) $\{x \in X : \|x - a\| < r\}$;
(3) $\{x \in X : r < \|x - a\| < s\}$.

Definition 1.3.16 ($\epsilon-$Ball, $\epsilon-$Neighborhood)**.** *The* **open $\epsilon-$ball ($\epsilon-$neighborhood)** *about a point x_0 in a normed linear space $(X, \|\cdot\|)$ is the set defined by*

$$B(x_0, \epsilon) := \{x \in X : \|x - x_0\| < \epsilon\}.$$

Proposition 1.3.17. *A subset U of a normed linear space X is open if and only if for each $x \in U$ there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq U$.*

Proof. Assume that U is open and, by contradiction, suppose that there exists $x \in U$ such that no such $\epsilon > 0$ exists with $B(x, \epsilon) \subseteq U$. Then for all $n \in \mathbb{N}$, there exists $x_n \notin U$ such that $x_n \in B(x, 1/n)$. Note that, since $\|x_n - x\| < 1/n$ for all $n \in \mathbb{N}$, the sequence $\{x_n\}_{n\in\mathbb{N}}$ converges to x. But $\{x_n\}_{n\in\mathbb{N}}$ is a sequence in U^C, and, since U^c is closed, we have $x \in U^C$, a contradiction.

Conversely, suppose that for all $x \in U$, there exists $\epsilon_x > 0$ such that $B(x, \epsilon_x) \subseteq U$. By De Morgan's Laws, observe

$$\bigcup_{x\in X} B(x, \epsilon_x) = \bigcup_{x\in X} \left(B(x, \epsilon_x)^C\right)^C = \left(\bigcap_{x\in X} B(x, \epsilon_x)^C\right),$$

so that the union $\bigcup_{x\in X} B(x, \epsilon_x)$ is open. It follows

$$U \subseteq \bigcup_{x\in X} B(x, \epsilon_x) \subseteq U,$$

and we see that evidently $U = \bigcup_{x\in X} B(x, \epsilon_x)$. Consequently U is open. □

Definition 1.3.18 (Topology)**.** *The collection of all open sets in a normed linear space X is called the* **topology** *of X, denoted by $\mathcal{T}$.*

The topology $\mathcal{T}$ for a normed linear space X satisfies the usual axioms of a topology, that is,

(1) $\emptyset, X \in \mathcal{T}$;
(2) If $U_1, U_2, \ldots, U_N \in \mathcal{T}$, then $\bigcap_{i=1}^N U_i \in \mathcal{T}$;
(3) If $\{U_\alpha\}_{\alpha\in A} \subseteq \mathcal{T}$, then $\bigcup_{\alpha\in A} U_\alpha \in \mathcal{T}$.

Definition 1.3.19 (Convergent Series)**.** *Let X be a normed linear space and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence in X. The series $\sum_{k=1}^\infty x_k$ is said to* **converge** *if the sequence of partial sums*

$$S_n := \sum_{k=1}^n x_k$$

converges.

Definition 1.3.20 (Absolute Convergence)**.** *A series $\sum_{k=1}^\infty x_k$ in a normed linear space is said to be* **absolutely convergent** *if the series of real numbers*

$$\sum_{k=1}^\infty \|x_k\|$$

is convergent.

An absolutely convergent series is convergent if the space is complete.

Theorem 1.3.21. *If a series in a Banach space converges absolutely, then all rearrangements of the series converge to a common value.*

Proof. Let $\sum_{k=1}^\infty x_k$ be an absolutely convergent series and $\sum_{j=1}^\infty x_{k_j}$ a permutation of it. Put

$$x := \sum_{k=1}^\infty x_k, \quad S_n := \sum_{k=1}^n x_k, \quad T_n := \sum_{j=1}^n x_{k_j}, \quad M := \sum_{k=1}^\infty \|x_k\|.$$

Clearly $\sum_{k=1}^{n}\|x_{k_j}\| \le M$, and thus $\sum_{k=1}^{\infty}\|x_{k_j}\| \le \infty$. Therefore $\sum_{k=1}^{\infty} x_{k_j}$ converges absolutely and hence converges by the completeness of the Banach space.

Put $y := \sum_{k=1}^{\infty} x_{k_j}$ and fix $\epsilon > 0$. Choose N so large that $\sum_{k=N}^{\infty}\|x_k\| < \epsilon$ and $\|S_m - x\| < \epsilon$ whenever $m \ge N$. Choose r so that $\|T_r - y\| < \epsilon$ and $\{1, \dots, N\} \subset \{k_1, \dots, k_r\}$. Choose m so large that $\{k_1, \dots, k_r\} \subset \{1, \dots, m\}$. Then $m \ge N$ and

$$\|S_m - T_r\| = \|(x_1 + x_2 + \cdots + x_m) - (x_{k_1} + x_{k_1} + \cdots + x_{k_r})\| \le \sum_{j=N+1}^{m}\|x_j\| < \epsilon.$$

Hence

$$\|x - y\| \le \|x - S_m\| + \|S_m - T_r\| + \|T_r - y\| < \epsilon + \epsilon + \epsilon = 3\epsilon.$$

This completes the proof. ☐

Theorem 1.3.22 (Riemann's Theorem). *If a series of real numbers is convergent but not absolutely, then for every real number, some rearrangement of the series converges to that real number.*

Proof. Let the series $\sum x_n$ converge but not absolutely. Then $\lim_{n\to\infty} x_n = 0$ and

$$\sum_{x_n>0} x_n - \sum_{x_n<0} x_n = \sum |x_n| = \infty.$$

Since the series $\sum x_n$ converges, the two series on the LHS must diverge to $+\infty$ and $-\infty$, respectively. Now let $r \in \mathbb{R}$ be any real number. Select positive terms (in order) from the series until their sum exceeds r. Now add negative terms (in order) until the new partial sum is less than r. Continue in this fashion. Since $\lim_{n\to\infty} x_n = 0$, the partial sums created in this process differ from r by quantities that tend to zero as $n \to \infty$. ☐

1.4. More About Compactness.

We first recall the Heine–Borel Theorem.

Theorem 1.4.1 (Heine–Borel Theorem). *A subset $E \subseteq \mathbb{R}$ is compact if and only if E is closed and bounded.*

We show that the Heine–Borel theorem is true for any normed linear space if and only if the space is finite–dimensional.

Lemma 1.4.2. *In the space $(\mathbb{R}^n, \|\cdot\|_\infty)$ each closed ball $\{x \in \mathbb{R}^n : \|x\|_\infty \le C\}$ is compact.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ be any sequence of points in $\mathbb{R}^n$ satisfying $\|x_n\|_\infty \le C$ for all $n \in \mathbb{N}$. Then the components obey the inequality

$$-C \le x_n(k) \le C$$

for all $n \in \mathbb{N}$ and each $k = 1, 2, \dots, n$. By the compactness of the interval $[-C, C]$, there exists an increasing sequence $I_1 := \{n_{1_i}\}_{i\in\mathbb{N}} \subseteq \mathbb{N}$ such that $\lim_{i\in I_1} x_i(1)$ exists. Similarly, there exists another increasing sequence $I_2 := \{n_{2_i}\}_{i\in\mathbb{N}} \subseteq I_1$ such that $\lim_{i\in I_2} x_i(2)$ exists. Moreover, since $I_2 \subseteq I_2$, $\lim_{i\in I_2} x_i(1)$ exists also.

Continuing in this fashion, we obtain at the $n-$th step an increasing sequence $I_n \subseteq I_{n-1} \subseteq \cdots \subseteq \mathbb{N}$ such that $\lim_{i\in I_n} x_i(k)$ exists for each $k = 1, 2, \dots, n$. Denote the limit by x^*. We have thus defined a vector x^* such that $\|x_k - x^*\|_\infty \to 0$ as k runs through the sequence of integers I_n. ☐

Lemma 1.4.3. *A closed subset of a compact set in a normed linear space is compact.*

Proof. Let F be a closed subset of a compact set K in a normed linear space X and let $\{x_n\}_{n\in\mathbb{N}}$ be a sequence of points in F. Since $F \subseteq K$ with K compact, there exists a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ of $\{x_n\}_{n\in\mathbb{N}}$ that converges to a point $x \in K$. Since F is closed, we must have $x \in F$ also, which completes the proof. □

Definition 1.4.4 (Bounded Subset)**.** *A subset S of a normed linear space is said to be* **bounded** *if there exists $C \geq 0$ such that $\|x\| \leq C$ for all $C \in S$. Equivalently,* $\sup_{x\in S} \|x\| < \infty$.

Theorem 1.4.5. *In any finite–dimensional normed linear space $(X, \|\cdot\|)$, every closed and bounded set is compact.*

Proof. Let $(X, \|\cdot\|)$ be a finite–dimensional normed linear space. There exists a basis for X, say $\{x_1, x_2, \ldots, x_n\}$. Define a mapping $T : \mathbb{R}^n \to X$ by the equation

$$T(a) := \sum_{k=1}^{n} a_k x_k, \quad a = [a_1, a_2, \ldots, a_n]^\top \in \mathbb{R}^n.$$

Assigning the norm $\|\cdot\|_\infty$ to $\mathbb{R}^n$, we see that T is continuous, for observe that

$$\begin{aligned}
\|T(a) - T(b)\| &= \left\| \sum_{k=1}^{n} (a_k - b_k) x_k \right\| \\
&\leq \sum_{k=1}^{n} \|(a_k - b_k) x_k\| \\
&= \sum_{k=1}^{n} |a_k - b_k| \|x_k\| \\
&\leq \left\{ \max_{k=1,2,\ldots,n} |a_k - b_k| \right\} \cdot \sum_{k=1}^{n} \|x_k\| \\
&= \|a - b\|_\infty \sum_{k=1}^{\infty} \|x_k\|.
\end{aligned}$$

Now let $F \subseteq X$ be a closed and bounded subset of X. Put $E := T^{-1}(F)$. Since E is the continuous preimage of a closed set, it follows that E is closed in $\mathbb{R}^n$ (1.3.12). Moreover, Since $F = T(E)$, it suffices to show that E is compact by (1.3.4). By (1.4.2) and (1.4.3), we need only show that E is bounded.

Define

$$\beta := \inf\{\|T(a)\| : \|a\|_\infty = 1\}.$$

Since the unit circle is compact, β is the infimum of a continuous map on a compact set. Hence β is attained by T at some point $b \in \mathbb{R}^n$. Thus $\|b\|_\infty = 1$ and

$$\beta = \|T(b)\| = \left\| \sum_{k=1}^{n} b_k x_k \right\|.$$

Since the points x_k, $k = 1, 2, \ldots, n$, constitute a linearly independent set, and since $b \neq 0$, we must have $T(b) \neq 0$ and that $\beta > 0$. Since F is bounded by the assumption, there exists a constant $C \geq 0$ such that $\|x\| \leq C$ for all $x \in F$.

Now let $a \in E$ be arbitrary. First note that if $a = 0$, then clearly $\|a\|_\infty \leq C/\beta$. Now if $a \neq 0$, then $a/\|a\|_\infty \in \{x \in \mathbb{R}^n : \|x\|_\infty = 1\}$, and we observe

$$\begin{aligned}\beta \leq \|T(a/\|a\|_\infty)\| &= \left\| \sum_{k=1}^{n} \frac{a_k}{\|a\|_\infty} x_k \right\| \\ &= \frac{1}{\|a\|_\infty} \left\| \sum_{k=1}^{n} a_k x_k \right\| \\ &= \frac{1}{\|a\|_\infty} \|T(a)\| \\ &\leq \frac{1}{\|a\|_\infty} C.\end{aligned}$$

Thus for any $a \in E$,

$$\|a\|_\infty \leq \frac{C}{\beta}.$$

It follows that E is bounded.

Finally, we see that E is a closed subset of the compact set

$$\{x \in \mathbb{R}^n : \|x\|_\infty \leq C/\beta\}.$$

Thus E is compact, and consequently, so is $F = T(E)$. This completes the proof. □

Corollary 1.4.6. *Every finite–dimensional normed linear space is complete.*

Proof. Let $\{x_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in a finite–dimensional normed linear space $(X, \|\cdot\|)$. We first show that $\{x_n\}_{n\in\mathbb{N}}$ us bounded. Choose a positive integer M such that $\|x_n - x_m\| \leq 1$ for all $n, m \geq M$. Then for all $n \geq M$,

$$\|x_n\| \leq \|x_n - x_m\| + \|x_m\| \leq 1 + \|x_m\|.$$

Hence for all $n \in \mathbb{N}$,

$$\|x_n\| \leq \|x_1\| + \|x_2\| + \cdots + \|x_m\| + 1.$$

Put $C := \|x_1\| + \|x_2\| + \cdots + \|x_m\| + 1$. Then $\{x_n\}_{n\in\mathbb{N}}$ is bounded by C.

Fix $\epsilon > 0$. Since $\{x \in X : \|x\| \leq C\}$ is compact, $\{x_n\}_{n\in\mathbb{N}}$ has a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ converging to a point $x \in X$. Choose a positive integer N_1 so large that for all $n \geq N_1$, we have $\|x_{n_{N_1}} - x\| < \epsilon/2$. Likewise, since $\{x_n\}_{n\in\mathbb{N}}$ is Cauchy, there exists a positive integer N_2 such that for all $n, m \geq N_2$, we have $\|x_n - x_m\| < \epsilon/2$. But then since $n_m \geq m$, we have in this situation $\|x_n - x_{n_m}\| < \epsilon/2$ as well.

Choose $N := \max\{N_1, N_2\}$. Then for all $n \geq N$, it follows

$$\|x_n - x\| \leq \|x_n - x_{n_N}\| + \|x_{n_N} - x\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Thus $\{x_n\}_{n\in\mathbb{N}}$ converges, and it follows that X is complete. □

We recall the following definition from linear algebra.

Definition 1.4.7 (Subspace)**.** *A subset Y of a normed linear space $(X, \|\cdot\|)$ is called a* **subspace** *of X if $(Y, \|\cdot\|)$ is a normed linear space.*

Corollary 1.4.8. *Every finite–dimensional subspace in a normed linear space is closed.*

Proof. Let Y be a finite–dimensional subspace of a normed linear space X and let $\{y_n\}_{n\in\mathbb{N}}$ be a sequence converging to a point $y \in X$. Since $\{y_n\}_{n\in\mathbb{N}}$ converges, clearly $\{y_n\}_{n\in\mathbb{N}}$ is Cauchy. But since Y is finite–dimensional, Y is complete (1.4.6), and thus $\{y_n\}_{n\in\mathbb{N}}$ converges to a point in Y, so that $y \in Y$. This completes the proof. □

Theorem 1.4.9 (Riesz's Lemma)**.** *If U is a closed and proper subspace in a normed linear space $(X, \|\cdot\|)$, and $\lambda \in \mathbb{R}$ is such that $0 < \lambda < 1$, then there exists a point $x \in X$ such that $\|x\| = 1$ and dist$(x, U) > \lambda$.*

Proof. Since U is proper, there exists a point $z \in U^C$. Also, since U is closed, $\operatorname{dist}(z, U) > 0$. By definition of $\operatorname{dist}(z, U)$, there is an element $u \in U$ such that

$$\lambda\|z - u\| < \operatorname{dist}(z, U),$$

for if not, then

$$\operatorname{dist}(z, U) = \inf_{u\in U} \|z - u\| \leq \lambda\|z - u\| < \|z - u\|$$

for all $u \in U$, a contradiction. Put $x := (z - u)/\|z - u\|$. Clearly $\|x\| = 1$. Moreover, we have

$$d(x, U) = \frac{\operatorname{dist}(z - u, U)}{\|z - u\|} = \frac{\operatorname{dist}(z, U)}{\|z - u\|} > \lambda.$$

□

Theorem 1.4.10. *If the unit ball in a normed linear space is compact, then the space has finite dimension.*

Proof. Let X be a normed linear space with a compact unit ball, and by contradiction, suppose that X is not finite–dimensional. We construct a sequence inductively as follows.

Let $x_1 \in X$ be any point such that $\|x_1\| = 1$. Given $x_1, \ldots, x_{n-1}$, denote by $U_{n-1} := \operatorname{span}\{x_1, x_2, \ldots, x_{n-1}\}$. Since every finite–dimensional subspace in a normed linear space is closed (1.4.8), U_{n-1} is closed. By Riesz's Lemma (1.4.9), we may select $x_n \in X$ such that $\|x_n\| = 1$ and $\operatorname{dist}(x_n, U_{n-1}) > 1/2$. Then $\|x_n - x_m\| > 1/2$ whenever $n > m$. This sequence cannot have any convergent subsequence, a contradiction to the unit ball being compact. □

Theorem 1.4.11 (Characterization of Finite–Dimensional Compact Spaces)**.** *A normed linear space is finite–dimensional if and only if its unit ball is compact.*

Proof. The proof follows immediately from (1.4.5) and (1.4.10). □

In any normed linear space, any compact set is necessarily closed and bounded. In finite–dimensional spaces, these two conditions are also *sufficient* for compactness. This may not hold in infinite–dimensional spaces, however.

1.5. **Linear Transformations.**

Definition 1.5.1 (Linear Transformation, Linear Operator)**.** *Let X and Y be two linear spaces over the same scalar field K. A mapping $f : X \to Y$ is called a* **linear transformation (linear operator)** *if*

$$f(\alpha u + \beta v) = \alpha f(u) + \beta f(v)$$

for all $\alpha, \beta \in K$ and $u, v \in X$.

Definition 1.5.2 (Linear Functional)**.** *Let X be a linear space over the scalar field K. A* **linear functional** *is a linear operator $f : X \to K$ from X to K.*

Note that by taking $\alpha = \beta = 0$ in the definition, we see immediately that a linear transformation T must have the property $T(0) = 0$.

Example 1.5.3. *If $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$, then each linear map of X into Y is of the form $f(x) = y$,*

$$y_i = \sum_{k=1}^{n} a_{i_k} x_k, \quad 1 \le i \le m,$$

where the a_{ik} are certain real numbers that form an $m \times n$ matrix.

Example 1.5.4. *Let $X = \mathcal{C}[0,1]$ and $Y = \mathbb{R}$. The mapping $f(x) = \int_0^1 x(t)\ dt$ is a linear functional.*

Example 1.5.5. *Let $X = \mathcal{C}^n[0,1]$ and let $a_0, a_1, \ldots, a_n$ be fixed elements of $\mathbb{R}$. Then a linear operator D is defined by*

$$Dx = \sum_{i=0}^{n} a_i x^{(i)}.$$

We call the operator D a **differential operator***.*

Example 1.5.6. *Let $X = \mathcal{C}[0,1] = Y$. Let k be a continuous function on $[0,1] \times [0,1]$. Define K by*

$$(Kx)(s) = \int_0^1 k(s,t) x(t)\ dt.$$

Then K is a linear operator, called a **linear integral operator***.*

Example 1.5.7. *Let X be the set of all bounded continuous functions on $\mathbb{R}^+ := \{t \in \mathbb{R} : t \ge 0\}$. Put*

$$(\mathcal{L}x)(s) = \int_0^\infty e^{-st} x(t)\ dt.$$

Then $\mathcal{L}$ is a linear operator, called the **Laplace transform***.*

Example 1.5.8. *Let X be the set of all continuous functions on $\mathbb{R}$ with $\int_{-\infty}^{\infty} |x(t)|\ dt < \infty$. Define*

$$(\mathcal{F}x)(s) = \int_{-\infty}^{\infty} e^{-2\pi i s t} x(t)\ dt.$$

Then $\mathcal{F}$ is a linear operator, the **Fourier transform***.*

Theorem 1.5.9. *Let X and Y be normed linear spaces. A linear operator $T : X \to Y$ is continuous if and only if T is continuous at zero.*

Proof. Let $T : X \to Y$ be a linear operator.

First, if T is continuous then clearly T is continuous at zero.

For the converse, suppose that T is continuous at zero. Then for every $\epsilon > 0$ there exists $\delta > 0$ such that for all $x \in X$ satisfying $\|x\|_X < \delta$, we have $\|T(x)\|_Y < \epsilon$. Now let $x, y \in X$ be such that $\|x - y\|_X < \delta$. Observe

$$\|T(x) - T(y)\|_Y = \|T(x - y)\|_Y < \epsilon.$$

Hence T is uniformly continuous and thus is continuous. □

Definition 1.5.10 (Bounded Linear Operator)**.** *A linear operator $T : X \to Y$ is called a* **bounded linear operator** *if T is bounded on the unit ball, that is,*

$$\sup\{\|T(x)\|_Y : \|x\|_X \leq 1\} < \infty.$$

Example 1.5.11. *Let $X = \mathcal{C}^1[0,1]$ and let X have the infinity norm $\|x\|_\infty := \sup_{t\in[0,1]} |x(t)|$. Let f be the linear functional defined by $f(x) = x'(1)$. Then this functional f is unbounded. To see this, consider the vectors $x_n(t) = t^n$ for each $n \in \mathbb{N}$.*

Theorem 1.5.12 (Continuity and Boundedness)**.** *A linear operator $T : X \to Y$ between two normed linear spaces is continuous if and only if it is bounded.*

Proof. Let $T : X \to Y$ be a linear operator.

First, assume that T is continuous. Choosing $\epsilon = 1$, there exists $\delta > 0$ such that for all $x \in X$ satisfying $\|x\|_X \leq \delta$, we have $\|T(x)\|_Y \leq 1$. Now let $x \in X$ be such that $\|x\|_X \leq 1$. Since $T(x) = \frac{1}{\delta}T(\delta x)$ and δx is such that $\|\delta x\|_x = |\delta|\|x\|_X \leq |\delta| = \delta$, it follows

$$\|T(x)\|_Y = \frac{1}{\delta}\|T(\delta x)\| \leq \frac{1}{\delta},$$

which shows that T is bounded.

Conversely, suppose that there exists $M > 0$ such that $\|T(x)\|_Y \leq M$ for all $x \in X$ such that $\|x\|_X \leq 1$. It suffices to show that T is continuous at zero. Fix $\epsilon > 0$. Then for all $x \in X$ such that $\|x\|_X < \frac{\epsilon}{2M}$, it follows

$$\|T(x)\|_Y = \left\|\frac{\epsilon}{2M}T\left(\frac{2Mx}{\epsilon}\right)\right\|_Y = \left|\frac{\epsilon}{2M}\right|\left\|T\left(\frac{2Mx}{\epsilon}\right)\right\|_Y \leq \left(\frac{\epsilon}{2M}\right)M < \epsilon,$$

which shows continuity at zero. This completes the proof. □

Proposition 1.5.13. *Let $\mathcal{L}(X,Y)$ be the family of all bounded linear operators from a normed linear space X into a normed linear space Y. For any $T \in \mathcal{L}(X,Y)$, introduce the operator norm*

$$\|T\| := \sup\{\|T(x)\|_Y := \|x\|_X \leq 1\}.$$

Then $\mathcal{L}(X,Y)$ is a normed linear space with respect to $\|\cdot\|$.

Proof. It is clear from the linearity of the elements of $\mathcal{L}(X,Y)$ that $\mathcal{L}(X,Y)$ forms a linear space. We show that $\|\cdot\|$ defines a norm on $\mathcal{L}(X,Y)$.

First, if $T \neq 0$, then there exists $x_0 \in X$ such that $\|x_0\|_X \leq 1$ and $T(x_0) \neq 0$. Hence,

$$\|T\| = \sup\{\|T(x)\|_Y : \|x\|_X \leq 1\} \geq \|T(x_0)\|_Y > 0,$$

by definition of $\|\cdot\|_Y$.

Next, for any $\alpha \in \mathbb{R}$, we have

$$\|\alpha T\| = \sup\{\|\alpha T(x)\|_Y : \|x\|_X \leq 1\} = |\alpha| \sup\{\|T(x)\|_Y : \|x\|_X \leq 1\} = |\alpha|\|T\|.$$

Finally, let $T_1, T_2 \in \mathcal{L}(X,Y)$. Observe

$$\begin{aligned}\|T_1 + T_2\| &= \sup\{\|(T_1+T_2)(x)\|_Y : \|x\|_X \leq 1\} \\ &\leq \sup\{\|T_1(x)\|_Y + \|T_2(x)\|_Y : \|x\|_X \leq 1\} \\ &\leq \sup\{\|T_1(x)\|_Y : \|x\|_X \leq 1\} + \sup\{\|T_2(x)\|_Y : \|x\|_X \leq 1\} \\ &= \|T_1\| + \|T_2\|.\end{aligned}$$

This completes the proof. $\square$

Lemma 1.5.14. *For any $T \in \mathcal{L}(X,Y)$,*

(1) $\|T(x)\|_Y \leq \|T\|\|x\|_X$;
(2) $\|T\| = \sup\{\|T(x)\|_Y : \|x\|_X = 1\}$.

Proof. Let $T \in \mathcal{L}(X,Y)$.

To show (1), let $x \in X$ be arbitrary. Then $x/\|x\|_X$ is a vector with norm precisely one, and

$$\begin{aligned}\|T(x)\|_Y &= \left\| \|x\|_X T\left(\frac{x}{\|x\|_X}\right)\right\|_Y = \|x\|_X \left\|T\left(\frac{x}{\|x\|_X}\right)\right\|_Y \\ &\leq \|x\|_X \left(\sup\{\|T(x)\|_Y : \|x\| \leq 1\}\right) \\ &= \|T\|\|x\|_X,\end{aligned}$$

which shows (1).

We now show (2). Clearly $\|T\| \geq \sup\{\|T(x)\|_Y : \|x\|_X = 1\}$. On the other hand, for all $x \in X$ such that $\|x\|_X \leq 1$,

$$\sup\{\|T(x)\|_Y : \|x\|_X = 1\} \geq \left\|T\left(\frac{x}{\|x\|_X}\right)\right\|_Y = \frac{\|T(x)\|_Y}{\|x\|_X} \geq \|T(x)\|_Y.$$

This is for all $\|x\|_X \leq 1$, so taking the supremum on the RHS gives

$$\sup\{\|T(x)\|_Y : \|x\|_X = 1\} \geq \sup\{\|T(x)\|_Y : \|x\|_X \leq 1\}.$$

This completes the proof. $\square$

We recall the following definition from linear algebra.

Definition 1.5.15 (Kernel)**.** *Let $T : X \to Y$ be a linear transformation. The* **kernel** *of T, denoted* $\ker(T)$, *is defined by*

$$\ker(T) := \{x \in X : T(x) = 0\}.$$

Theorem 1.5.16. *A linear functional $f : X \to \mathbb{R}$ on a normed linear space X is continuous if and only if its kernel is closed.*

Proof. Let $f : X \to \mathbb{R}$ be a linear functional.

Assume that f is continuous. Since $\{0\}$ is closed in $\mathbb{R}$, it follows by the continuity of f that $\ker(f) = f^{-1}(0)$ is closed in X.

Now suppose that $\ker(f)$ is closed. By contradiction, suppose further that f is discontinuous. Then f is unbounded, so that there exists a sequence $\{x_n\}_{n\in\mathbb{N}}$ in $\ker(f)$ such that $\|x_n\| \leq 1$ but $f(x_n) \to \infty$ as $n \to \infty$.

If $(\ker(f))^C = \emptyset$, then $f \equiv 0$ and thus continuous. Otherwise, choose $x \notin \ker(f)$. Define a sequence $\{y_n\}_{n\in\mathbb{N}}$ by

$$y_n := x - \frac{f(x)}{f(x_n)} x_n$$

for each $n \in \mathbb{N}$. Then each y_n lies in the kernel of f, for

$$f(y_n) = f\left(x - \frac{f(x)}{f(x_n)} x_n\right) = f(x) - \frac{f(x)}{f(x_n)} f(x_n) = 0$$

for every $n \in \mathbb{N}$. Moreover, the sequence $\{y_n\}_{n\in\mathbb{N}}$ converges to x, for we see that

$$\|y_n - x\| = \left\|\left(x - \frac{f(x)}{f(x_n)} x_n\right) - x\right\| = \left\|\frac{f(x)}{f(x_n)} x_n\right\| \to 0 \quad \text{as} \quad n \to \infty,$$

since $f(x_n) \to \infty$ as $n \to \infty$. But since the kernel of f is closed, we have $x \in \ker(f)$, a contradiction. □

Corollary 1.5.17. *Every linear functional on a finite–dimensional normed linear space is continuous.*

Proof. Let $f : X \to \mathbb{R}$ be a linear functional with X a finite–dimensional normed linear space. Then $\ker(f)$ is a subspace of X. Since every subspace of a finite–dimensional normed linear space is closed (1.4.8), $\ker(f)$ is closed. Then, since a linear functional on a normed space is continuous if and only if its kernel is closed (1.5.16), we have that f is continuous. □

Corollary 1.5.18. *Every linear transformation from a finite–dimensional normed linear space to another normed linear space is continuous.*

Proof. Let $T : X \to Y$ be a linear transformation with X finite–dimensional. There exists a basis for X, say $\{b_1, b_2, \ldots, b_n\}$. Choose $x \in X$. We may express x as a unique linear combination of the basis elements $b_1, b_2, \ldots, b_n$. That is, there exists $\lambda_k(x)$, $k = 1, 2, \ldots, n$, such that

$$x = \sum_{k=1}^{n} \lambda_k(x) b_k.$$

We show that the λ_k, $k = 1, 2, \ldots, n$, are linear functionals.

Let $u \in X$ be arbitrary. Note

$$\alpha x + \beta u = \sum_{k=1}^{n} \lambda_k(\alpha x + \beta u) b_k.$$

On the other hand,

$$\alpha x + \beta u = \alpha \sum_{k=1}^{n} \lambda_k(x) b_k + \beta \sum_{k=1}^{n} \lambda_k(u) b_k = \sum_{k=1}^{n} (\alpha\lambda_k(x)) b_k + \sum_{k=1}^{n} (\beta\lambda_k(u)) b_k$$

$$= \sum_{k=1}^{n} (\alpha\lambda_k(x) + \beta\lambda_k(u)) b_k.$$

By uniqueness of the coefficients, we conclude that

$$\lambda_k(\alpha x + \beta u) = \alpha\lambda_k(x) + \beta\lambda_k(x)$$

for all $k = 1, 2, \dots, n$, which shows that each λ_k, $k = 1, 2, \dots, n$, is a linear functional.

Since every linear functional on a finite–dimensional normed linear space is continuous (1.5.17), it follows that each λ_k, $k = 1, 2, \dots, n$, is continuous. Finally,

$$T(x) = T\left(\sum_{k=1}^{n} \lambda_k(x) b_k\right) = \sum_{k=1}^{n} \lambda_k(x) T(b_k)$$

is clearly continuous. □

Corollary 1.5.19 (Equivalence of Norms)**.** *All norms on a finite–dimensional normed linear space X are equivalent, in the sense that for each pair of norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on X, there exist $m, M > 0$ such that*

$$m\|x\|_\alpha \le \|x\|_\beta \le M\|x\|_\alpha$$

for all $x \in X$.

Proof. Let X be a finite–dimensional normed linear space and let $\|\cdot\|_\alpha$, $\|\cdot\|_\beta$ be two norms on X. Since every linear transformation from a finite–dimensional normed space to another normed space is continuous (1.5.18), the identity operator $\mathbb{1}_X^{(\alpha)} : (X, \|\cdot\|_\alpha) \to (X, \|\cdot\|_\beta)$ defined by $\mathbb{1}_X^{(\alpha)}(x) = x$ for all $x \in X$ is continuous. Thus $\mathbb{1}_X$ is bounded, so that there exists $M > 0$ such that

$$\|x\|_\beta = \|\mathbb{1}_X^{(\alpha)}(x)\|_\beta \le M\|x\|_\alpha.$$

On the other hand, $\mathbb{1}_X^{(\beta)} : (X, \|\cdot\|_\beta) \to (X, \|\cdot\|_\alpha)$ is continuous. Similarly, there exists $k > 0$ such that

$$\|x\|_\alpha = \|\mathbb{1}_X^{(\beta)}(x)\|_\alpha \le k\|x\|_\beta.$$

Taking $m := 1/k$, we have

$$m\|x\|_\alpha \le \|x\|_\beta \le M\|x\|_\alpha$$

for all $x \in X$. This completes the proof. □

Recall that $\mathcal{L}(X, Y)$ denotes the set of all linear bounded operators from X to Y. The space $\mathcal{L}(X, Y)$ has a natural structure. More specifically, we define

$$(\alpha A + \beta B)(x) = \alpha(Ax) + \beta(Bx),$$

$$\|A\| = \sup\{\|Ax\|_Y : x \in X, \|x\|_X \le 1\},$$

where $A, B \in \mathcal{L}(X, Y)$ and $\alpha, \beta \in \mathbb{R}$.

Theorem 1.5.20 ($\mathcal{L}(X, Y)$ is a Banach Space)**.** *If X is a normed linear space and Y is a Banach space, then $\mathcal{L}(X, Y)$ is a Banach space.*

Proof. We have already shown that $\mathcal{L}(X, Y)$ is a normed linear space with respect to the operator norm $\|\cdot\|$ (1.5.13). It remains to show that $\mathcal{L}(X, Y)$ is complete.

Let $\{A_n\}_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{L}(X, Y)$. Fix $x \in X$. We have

$$\|A_n(x) - A_m(x)\|_Y = \|(A_n - A_m)(x)\|_Y \leq \|A_n - A_m\|\|x\|_X,$$

which shows that $\{A_n(x)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in Y for any $x \in X$. Since Y is a Banach space, we can define the pointwise limit $Ax = \lim_{n\to\infty} A_n x$. We show that $A \in \mathcal{L}(X, Y)$ and $\|A_n - A\| \to 0$ as $n \to \infty$.

We first show that $A \in \mathcal{L}(X, Y)$. Since each A_n, $n \in \mathbb{N}$ is linear, we have

$$A_n(\alpha x + \beta y) = \alpha A_n x + \beta A_n y$$

for all $x, y \in X$ and $\alpha, \beta \in \mathbb{R}$. Letting $n \to \infty$, we obtain

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay,$$

which establishes that A is linear. To see that A is bounded, first recall that Cauchy sequences are bounded. Thus there exists $M \geq 0$ such that $\|A_n\| \leq M$ for all $n \in \mathbb{N}$. Thus

$$\|A_n x\|_Y \leq \|A_n\|\|x\|_X \leq M\|x\|_X$$

for all $x \in X$. Letting $n \to \infty$, we have $\|Ax\| \leq M\|x\|_X$ for all $x \in X$, so that A is bounded. This shows that $A \in \mathcal{L}(X, Y)$.

Lastly, we show that $\{A_n\}_{n\in\mathbb{N}}$ converges to A in norm. Fix $\epsilon > 0$. Since $\{A_n\}_{n\in\mathbb{N}}$ is Cauchy, there exists a positive integer N such that for all $m, n \geq N$, we have $\|A_m - A_n\| < \epsilon$. Then

$$\|A_m x - A_n x\|_Y \leq \|A_m - A_n\|\|x\|_X \leq \|A_m - A_n\| < \epsilon$$

for all $x \in X$ such that $\|x\|_X \leq 1$. Letting $n \to \infty$, we have for all $\|x\|_X \leq 1$ that

$$\|A_m x - Ax\|_Y \leq \|A_m - A\|\|x\|_X \leq \|A_m - A\| < \epsilon.$$

This completes the proof. □

For any $A, B \in \mathcal{L}(X, Y)$, the composition of A and B is generally written AB rather than $A \circ B$. That is, $(AB)x = A(Bx)$, and AB is linear. We denote the identity operator by I.

Theorem 1.5.21 (The Neumann Theorem)**.** *Let $A : X \to X$ be a bounded linear operator on a Banach space X. If $\|A\| < 1$, then $I - A$ is invertible, and*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Proof. Put $B_n : \sum_{k=0}^{n} A^k$. Then each B_n is itself a bounded linear operator, and $\{B_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence, for if $n > m$, we observe that

$$\|B_n - B_m\| = \left\|\sum_{k=0}^{n} A^k - \sum_{k=0}^{m} A^k\right\| = \left\|\sum_{k=m+1}^{n} A^k\right\|$$

$$= \sum_{k=m+1}^{n} \|A^k\| \leq \sum_{k=m}^{\infty} \|A\|^k = \|A\|^m \sum_{k=0}^{\infty} = \frac{\|A\|^m}{1 - \|A\|}.$$

Since $\|A\| < 1$ by the assumption, taking m sufficiently large guarantees the Cauchy criterion.

Since the space of all linear bounded operators on a Banach space is itself a Banach space (1.5.20), the sequence $\{B_n\}_{n\in\mathbb{N}}$ converges to a linear bounded operator $B \in \mathcal{L}(X, X)$. We have

$$(I - A)B_n = B_n - AB_n = \sum_{k=0}^{n} A^n - \sum_{k=0}^{n} A^{k+1} = \sum_{k=0}^{n} A^k - \sum_{k=1}^{n+1} = A^0 - A^{n+1} = I - A^{n+1}.$$

Taking $n \to \infty$, we obtain $(I - A)B = I$. Similarly, $B(I - A) = I$. Consequently $I - A$ is invertible with

$$(I - A)^{-1} = B = \sum_{k=0}^{\infty} A^k.$$

□

1.6. **Zorn's Lemma, Hamel Bases, and the Hahn–Banach Theorem.** The following theorem is due to Gödel.

Theorem 1.6.1. *If a contradiction can be derived from the Zermelo–Fraenkel axioms of set theory (which include the axiom of choice), then a contradiction can be derived within the restricted set theory based on the Zermelo–Fraenkel axioms without the axiom of choice.*

Theorem 1.6.2 (Axiom of Choice)**.** *If A is a set and f a function on A such that $f(\alpha)$ is a nonempty set for each $\alpha \in A$, then f has a choice function. That is, there exists a function c on A such that $c(\alpha) \in f(\alpha)$ for all $\alpha \in A$.*

Example 1.6.3. *Let A be a finite set, $A := \{\alpha_1, \dots, \alpha_n\}$. For each i in $\{1, 2, \dots, n\}$, a nonempty set $f(\alpha_i)$ is given. In n steps, we can select elements $x_1 \in f(\alpha_1), x_2 \in f(\alpha_2), \dots, x_n \in f(\alpha_n)$. Define then $c(\alpha_i) = x_i$ for each $i = 1, 2, \dots, n$.*

Attempting the same construction for infinite sets $A = \mathbb{R}$, for instance, leads to difficulty. The axiom of choice asserts that such a choice function exists, however.

Definition 1.6.4 (Partially Ordered Set)**.** *A* **partially ordered set** *is a pair $(X, \prec)$ in which X is a set and $\prec$ is a relation on X such that*

(1) For all $x \in X$, $x \prec x$;
(2) If $x \prec y$ and $y \prec z$, then $x \prec z$.

Definition 1.6.5 (Totally Ordered Set)**.** *A* **totally ordered set** *is a partially ordered set $(X, \prec)$ in which for any two elements $x, y \in X$, either $x \prec y$ or $y \prec x$.*

Definition 1.6.6 (Upper Bound)**.** *Let X be a partially ordered set and A a subset of X. An* **upper bound** *for A is any point $x \in X$ such that $a \prec x$ for all $a \in A$.*

Example 1.6.7. *Let S be any set and denote by $\mathcal{P}(S)$ the power set of S. Order $\mathcal{P}(S)$ by the inclusion relation $\subseteq$. Then $(\mathcal{P}(S), \subseteq)$ is a partially ordered set, but is not totally ordered. An upper bound for any subset of $\mathcal{P}(S)$ is S.*

Example 1.6.8. *In $\mathbb{R}^2$ define $x \prec y$ to mean that $x_i \leq y_i$ for $i = 1$ and $i = 2$. This is a partial ordering but not a total ordering. Note that only the third quadrant has an upper bound.*

Example 1.6.9. *Let $\mathcal{F}$ be a family of functions. For $f, g \in \mathcal{F}$ we write $f \prec g$ if the following two conditions are satisfied:*

(1) $dom(f) \subseteq dom(g)$;
(2) $f(x) = g(x)$ for all $x \in dom(f)$.

In this situation we call g an extension of f.

Definition 1.6.10 (Maximal Element)**.** *An element M in a partially ordered set X is said to be a* **maximal element** *if every $x \in X$ that satisfies the condition $M \prec x$ also satisfies $x \prec M$.*

Theorem 1.6.11 (Zorn's Lemma)**.** *Let X be a partially ordered set. If each totally ordered subset of X has an upper bound, then X contains a maximal element.*

Definition 1.6.12 (Hamel Basis)**.** *Let X be a linear space. A subset H of X is called a* **Hamel basis** *if each point $x \in X$ has a unique expression as a finite linear combination of elements of H.*

Example 1.6.13. *Let X be the space of all polynomials defined on $\mathbb{R}$. A Hamel basis for X is given by the sequence $\{h_n\}_{n\in\mathbb{N}_0}^{\infty}$, where $h_n(t) := t^n$ for each $n \in \mathbb{N}_0$.*

Theorem 1.6.14 (Existence of Hamel Basis)**.** *Every nontrivial linear space has a Hamel Basis.*

Proof. Let X be a nontrivial linear space. To show that X has a Hamel basis, we first prove that X has a maximal linearly independent set, and then we show that any such set is necessarily a Hamel basis.

Consider the collection $\mathcal{C}$ of all linearly independent subsets of X, and partially order $\mathcal{C}$ by the inclusion relation, $\subseteq$. In order to apply Zorn's Lemma (1.6.11), we verify that every totally ordered subset in $\mathcal{C}$ has an upper bound. Note $\mathcal{C}$ is nonempty, since, for any $x \in X$, $\{x\}$ is a linearly independent subset of X. Let $T \in \mathcal{C}$ be a totally ordered subset. Consider $S^* \bigcup_{S\in T} S$. Clearly $S \subseteq S^*$ for all $S \in T$. We show that S^* is linearly independent. Suppose that $\sum_{i=1}^{n} a_i s_i = 0$ for some scalars a_i and for some distinct points $s_i \in S^*$. Each s_i belongs to some $S_i \in T$. Since T is a totally ordered subset and there are only finitely many s_i, one of the sets S_i, $i = 1, 2, \ldots, n$, say, S_j, contains all the others. Since S_j is linearly independent by the assumption, we conclude that $a_i = 0$ for each $i = 1, 2, \ldots, n$. This shows that S^* is linearly independent, from which it follows that every totally ordered subset in $\mathcal{C}$ has an upper bound. By Zorn's Lemma, the collection $\mathcal{C}$ has a maximal element, say H.

We show that H is a Hamel basis for X. Let $x \in X$ be arbitrary. By the maximality of H, either $H \cup \{x\}$ is linearly dependent or $H \cup \{x\} \subseteq H$, in which case $x \in H$. In either case, x is a linear combination of elements of H. Since H itself is linearly independent, it follows that this linear combination is unique. Hence, H is a Hamel basis for X.

This completes the proof. □

Definition 1.6.15 (Dominated)**.** *Let $f, p : X \to \mathbb{R}$ be real–valued functions. We say that f is* **dominated** *by p if for all $x \in X$, we have*

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

Theorem 1.6.16 (Hahn–Banach Theorem)**.** *Let X be a real linear space, and let $p : X \to \mathbb{R}$ be a function such that $p(x+y) \leq p(x)+p(y)$ and $p(\lambda x) = \lambda p(x)$ if $\lambda \geq 0$. Then any linear functional defined on a subspace of X and dominated by p has an extension that is linear, defined on X, and dominated by p.*

Proof. Let X_0 be a linear subspace of X and let $f : X_0 \to \mathbb{R}$ be a linear functional. If $X_0 = X$ we're done. Otherwise, let $Y \in X \setminus X_0$. To extend f to $X_0 \cup \operatorname{span}(y)$ it suffices to specify a value for $f(y)$, for the linearity of f implies that

$$f(x + \lambda y) = f(x) + \lambda f(y)$$

for all $x \in X_0$ and all $\lambda \in \mathbb{R}$. In order for this extension to be dominated by p, the value of $f(y)$ must be assigned in such a way that

$$f(x + \lambda y) = f(x) + \lambda f(y) \leq p(x + \lambda y)$$

for all $x \in X_0$ and all $\lambda \in \mathbb{R}$. If $\lambda = 0$, the inequality is trivial. If $\lambda > 0$, we may divide by λ to see that we must have

$$f\left(\frac{x}{\lambda}\right) + f(y) = \frac{1}{\lambda} f(x) + f(y) \leq \frac{1}{\lambda} p(x + \lambda y) = p\left(\frac{x}{\lambda} + y\right)$$

for all $x \in X_0$, or

$$f(x_1) + f(y) \leq p(x_1 + y)$$

for all $x_1 \in X_0$. If $\lambda < 0$, dividing by λ and exchanging inequalities gives

$$f\left(\frac{x}{\lambda}\right) + f(y) \geq p\left(\frac{x}{\lambda} + y\right) = -p\left(-\frac{x}{\lambda} - y\right)$$

for all $x \in X_0$, or

$$f(x_2) + f(y) \geq -p(-x_2 - y)$$

for all $x_2 \in X_0$. Adding the above inequalities, we see that $f(y)$ must satisfy

$$-p(x_2 - y) - f(x_2) \leq f(y) \leq p(x_1 + y) - f(x_1)$$

for all $x_1, x_2 \in X_0$. In order to see that there is a number satisfying this inequality, we compute

$$\begin{aligned} f(x_1) - f(x_2) = f(x_1 - x_2) \leq p(x_1 - x_2) &= p((x_1 + y) - (x_2 + y)) \\ &\leq p(x_1 + y) + p(-x_2 - y). \end{aligned}$$

This completes the extension by one dimension.

We use Zorn's Lemma to show that f can be extended to X. Let $\mathcal{F}$ be the collection of all linear extensions of f that are dominated by p, and partially order $\mathcal{F}$ by the inclusion relation $\subseteq$. Note $\mathcal{F}$ is nonempty by the above argument. Recall that $h \subseteq g$ if and only if the domain of g contains the domain of h and $g(x) = h(x)$ on the domain of h. In order to use Zorn's Lemma, we must verify that each totally ordered subset in $\mathcal{F}$ has an upper bound. But this is true since the union of all the extensions in any totally ordered subset is an upper bound for this subset. By Zorn's Lemma, there exists a maximal element $\widetilde{f}$ in

$\mathcal{F}$. Then $\widetilde{f}$ is a linear functional that is an extension of f and is dominated by p. Finally, $\widetilde{f}$ must be defined on all of X, for if not, a further extension of f would be possible, as shown in the first part of the proof.

This completes the proof. □

Corollary 1.6.17. *Let f be a linear functional defined on a subspace Y in a normed linear space X and satisfying*

$$|f(y)| \le M\|y\|$$

for all $y \in Y$. Then f has a linear extension defined on all of X and satisfying the above inequality on X.

Proof. Apply the Hahn–Banach Theorem (1.6.16) with $p(x) := M\|x\|$. □

Corollary 1.6.18. *Let Y be a subspace in a normed linear space X. If $w \in X$ and $dist(x, Y) > 0$, then there exists a continuous linear functional f defined on X such that $f(y) = 0$ for all $y \in Y$, $f(w) = 1$, and $\|f\| = 1/dist(w, Y)$.*

Proof. Let V be the subspace generated by Y and w, that is, $V := \operatorname{span}\{Y, w\}$. Each element v of V has a unique representation $v = y+\lambda w$, with $y \in Y$ and $\lambda \in \mathbb{R}$. Let $f(v) = f(y+\lambda w) = \lambda$ for all $v \in V$. The norm of f on V is computed as follows:

$$\begin{aligned}\|f\| &= \sup_{v\neq 0} \frac{|f(y+\lambda w)|}{\|y+\lambda w\|} = \sup_{v\neq 0} \frac{|\lambda|}{\|y+\lambda w\|} = \sup_{v\neq 0} \frac{1}{\|\frac{y}{\lambda}+w\|} \\ &= \frac{1}{\inf_{y\in Y}\|y+w\|} = \frac{1}{\operatorname{dist}(w,Y)}.\end{aligned}$$

By the above corollary (1.6.17), we can extend f to all of X without an increase in norm. □

Corollary 1.6.19. *To each point w in a normed linear space there corresponds a continuous linear functional f such that $\|f\| = 1$ and $f(w) = \|w\|$.*

Proof. Fix $w \in X$ and let $Y = \{0\}$ in the above corollary (1.6.18). There exists a continuous linear functional $\phi : X \to \mathbb{R}$ such that $\phi(w) = 1$ and $\|\phi\| = 1/\|w\|$. Let $\psi(x) := \|w\|\phi(x)$. Then ψ is clearly a linear functional, and satisfies

$$\|\psi\| = 1, \quad \psi(w) = \|w\|.$$

□

Definition 1.6.20 (Dual Space, Adjoint)**.** *Let X be a normed linear space. The* **adjoint (dual space)** *of X is a normed space X^* consisting of all continuous linear functionals defined on X.*

Example 1.6.21. *Let $X = \mathbb{R}^n$ endowed with the max norm. Then X^* can be identified with the norm $\|\cdot\|_1$. To see this, recall that if $\phi \in X^*$, then $\phi(x) = \sum_{i=1}^n u_i x_i$ for a suitable $u \in \mathbb{R}^n$. Then*

$$\|\phi\| = \sup_{\|x\|_\infty \le 1} \left|\sum_{i=1}^n u_i x_i\right| = \sum_{i=1}^n |u_i| = \|u\|_1.$$

Example 1.6.22. *Let c_0 denote the Banach space of all real sequences that converge to zero, normed by putting $\|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n|$. Let ℓ_1 denote the Banach space of all real sequences u for which $\sum_{n=1}^\infty |u_n| < \infty$, normed by putting $\|u\|_1 := \sum_{n=1}^\infty |u_n|$. With each $u \in \ell_1$ we associate a functional $\phi_u \in c_0^*$ by means of the equation $\phi_u(x) = \sum_{n=1}^\infty u_n x_n$.*

Definition 1.6.23 (Isometric)**.** *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be normed linear spaces and $T : X \to Y$ a linear operator. We say that T is* **isometric** *if for all $x \in X$, we have*

$$\|Tx\|_Y = \|x\|_X.$$

Proposition 1.6.24. *For each $u \in \ell_1$, define the functional $\phi_u : c_0 \to \mathbb{R}$ by $\phi_u(x) = \sum_{n=1}^{\infty} u_n x_n$. The mapping $A : \ell_1 \to c_0^*$ defined by $Au = \phi_u$ is an isometric isomorphism between ℓ_1 and c_0^*.*

Proof. It is to be shown that for each $u \in \ell_1$, Au is linear and continuous on c_0. Then it is to be shown that A is linear, injective, surjective, and isometric.

First, that ϕ_u is well–defined follows from the absolute convergence of the series defining $\phi_u(x)$:

$$\|\phi_u\| \leq \sum_{n=1}^{\infty} |x_n u_n| = \sum_{n=1}^{\infty} |x_n||u_n| \leq \sum_{n=1}^{\infty} \|x\|_\infty |u_n| = \|x\|_\infty \sum_{n=1}^{\infty} |u_n| = \|x\|_\infty \|u\|_1.$$

The linearity of ϕ_u is obvious:

$$\begin{aligned} \phi_u(\alpha x + \beta y) &= \sum_{n=1}^{\infty} u_n(\alpha x_n + \beta y_n) = \sum_{n=1}^{\infty} \alpha u_n x_n + \beta u_n y_n = \alpha \sum_{n=1}^{\infty} u_n x_n + \beta \sum_{n=1}^{\infty} u_n y_n \\ &= \alpha \phi_u(x) + \beta \phi_u(y). \end{aligned}$$

The continuity of ϕ_u follows immediately from the boundedness of ϕ_u shown in the proof above that ϕ_u is well–defined. This shows Au is linear and continuous.

To see that A is isometric, observe

$$\|Au\| = \|\phi_u\| = \sup_{\|x\|_\infty \leq 1} \|x\|_\infty \|u\|_1 \leq \|u_1\|.$$

On the other hand, if $\epsilon > 0$ is given, we may select N so large that $\sum_{n=N+1}^{\infty} |u_n| < \epsilon$. Then we define x by putting $x_n := \operatorname{sgn}(u_n)$ for $n \leq N$, and by setting $x_n := 0$ for $n > N$. Clearly, $x \in c_0$ and $\|x\|_\infty = 1$. We have

$$\|\phi_u\| \geq \phi_u(x) = \sum_{n=1}^{N} x_n u_n = \sum_{n=1}^{N} |u_n| > \|u\|_1 - \epsilon.$$

Since ϵ was arbitrary, $\|\phi_u\| \geq \|u\|_1$, which shows that A is isometric, that is,

$$\|Au\| = \|\phi_u\| = \|u\|_1.$$

We now show that A is injective. Suppose that $Au_1 = Au_2$ for some $u_1, u_2 \in \ell_1$. Then $\phi_{u_1} = \phi_{u_2}$. Thus, for any $x \in c_0$, we have

$$\phi_{u_1}(x) - \phi_{u_2}(x) = \sum_{n=1}^{\infty} (u_1)_n x_n - \sum_{n=1}^{\infty} (u_2)_n x_n = \sum_{n=1}^{\infty} ((u_1)_n - (u_2)_n) x_n.$$

Since this is for all $x \in c_0$, we must have $u_1 = u_2$, which shows that A is injective.

Next, we show that A is surjective. Let $\psi \in c_0^*$. Let δ_n be the element of c_0 that has a 1 in the $n-$th coordinate and zeros elsewhere. Then for any $x \in c_0$,

$$x = \sum_{n=1}^{\infty} x_n \delta_n.$$

Since ψ is continuous and linear,

$$\psi(x) = \psi\left(\sum_{n=1}^{\infty} x_n \delta_n\right) = \sum_{n=1}^{\infty} x_n \psi(\delta_n).$$

Consequently, if we put $u_n := \psi(\delta_n)$, then $\psi(x) = \phi_u(x)$ for all $x \in c_0$ and $\psi = \phi_u$. To verify that $u \in \ell_1$, we define $y_n = \operatorname{sgn}(u_n)$ for $n \geq N$ and $y_n = 0$ for $n > N$. Then

$$\sum_{n=1}^{N} |u_n| = \sum_{n=1}^{N} y_n u_n = \psi(y) \leq \|\psi\| \|y\|_\infty = \|\psi\|,$$

by the boundedness of ψ. Thus $\|u\|_1 \leq \|\psi\|$.

Finally, the linearity of A follows from

$$\begin{aligned}\phi_{\alpha u + \beta v}(x) &= \sum_{n=1}^{\infty} (\alpha u_n + \beta v_n) x_n = \alpha \sum_{n=1}^{\infty} u_n x_n + \beta \sum_{n=1}^{\infty} v_n x_n = \alpha \phi_u(x) + \beta \phi_v(x) \\ &= (\alpha \phi_u + \beta \phi_v)(x).\end{aligned}$$

This completes the proof. □

Corollary 1.6.25. *For each x in a normed linear space X, we have*

$$\|x\| = \max\{|\phi(x)| : \phi \in X^*, \|\phi\| = 1\}.$$

Proof. If $\phi \in X^*$ and $\|\phi\| = 1$, then for all $x \in X$, we have

$$|\phi(x)| \leq \|\phi\| \|x\| = \|x\|.$$

Therefore

$$\sup\{|\phi(x)| : \phi \in X^*, \|\phi\| = 1\} \leq \|x\|.$$

For the reverse inequality, note first that it is trivial if $x = 0$. Otherwise, corollary (1.6.19) implies that there exists a continuous linear functional $\psi \in X^*$ such that $\|\psi\| = 1$ and $\psi(x) = \|x\|$. That is,

$$\|x\| = |\psi(x)| \leq \|\psi\| \|x\|.$$

Note that the supremum is attained. □

Definition 1.6.26 (Fundamental Subset)**.** *A subset V in a normed linear space X is said to be* **fundamental** *if the set of all linear combinations of elements of V is dense in X. That is, for all $x \in X$ and any $\epsilon > 0$, there exists $v_1, \ldots, v_n \in V$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that*

$$\left\| x - \sum_{i=1}^{n} \lambda_i v_i \right\| < \epsilon.$$

Note that we could also state that $\operatorname{dist}(x, \operatorname{span}(V)) = 0$ for all $x \in X$.

Example 1.6.27. *In the space $(\mathcal{C}[a,b], \|\cdot\|_\infty)$, an important fundamental set is the sequence of monomials*

$$u_0(t) := 0, \quad u_1(t) := t, \quad u_2(t) := t^2, \ldots$$

In fact, the Weierstrass Approximation Theorem establishes the fundamentality of this sequence. Thus for any $x \in \mathcal{C}[a,b]$ and any $\epsilon > 0$ there is a polynomial p for which $\|x - p\|_\infty < \epsilon$.

Definition 1.6.28 (Annihilator)**.** *If A is a subset of a normed linear space X, then the* **annihilator** *of A is the set*

$$A^\perp := \{\phi \in X^* : \forall a \in A, \phi(a) = 0\}.$$

Theorem 1.6.29. *A subset V of a normed linear space X is fundamental if and only if V has trivial annihilator.*

Proof. We first show that if $V^\perp = \{0\}$, then V is fundamental. By contrapositive, suppose that V is not fundamental. Denote by Y the closure of the span of V, $Y := \overline{\operatorname{span}(V)}$. If V is not fundamental, there exists a point $x \in X \setminus Y$. Then by corollary (1.6.18), there exists a continuous linear functional $\phi \in X^*$ such that $\phi(v) = 0$ for all $y \in Y$ and $\phi(x) = 1$. Since $\phi(y) = 0$ for all $y \in Y$, $\phi \in Y^\perp$. But, since $\phi(x) = 1$, $\phi \neq 0$, and we see that $V^\perp \neq \{0\}$. This proves the converse.

Now assume that V is fundamental, in which case $Y = X$. Then any element of $V^\perp$ annihilates the span of V as well as Y and thus X. Thus it must be the zero functional, that is, $V^\perp = \{0\}$. □

Theorem 1.6.30 (Adjoints are Banach Spaces)**.** *For any normed linear space X (not necessarily complete) its adjoint X^* is complete.*

Proof. For any normed linear space X, note that $X^* = \mathcal{L}(X, \mathbb{R})$. Since $\mathbb{R}$ is complete, the completeness of X^* follows from theorem (1.5.20). □

1.7. **The Baire Theorem and Uniform Boundedness.**

This section discusses important consequences of *completeness* in a normed linear space. That is, the following results distinguish Banach spaces from other normed linear spaces. These results should show why it is always nicer to be messing with a complete space.

Definition 1.7.1 (Dense)**.** *A subset F of a metric space X is said to be* **dense** *if its closure is the entire space, $\overline{F} = X$.*

Theorem 1.7.2 (Baire's Theorem)**.** *In any complete metric space, the intersection of a countable family of open dense sets is dense.*

Proof. Let $\{\mathcal{O}_n\}_{n=1}^\infty$ be a countable collection of open dense sets in a metric space X. In order to show that $\bigcap_{n=1}^\infty \mathcal{O}_n$ is dense, it suffices to prove that this set intersects an arbitrary nonempty open ball in X.

Fix $x \in X$ and $\epsilon > 0$ and define $S := B(x, \epsilon)$. Since $\mathcal{O}_1$ is open and dense, $S \cap \mathcal{O}_1$ is open and nonempty. Thus there exist $r_1 > 0$ and $x_1 \in S \cap \mathcal{O}_1$ such that $S_1 := B(x_1, r_1) \subseteq S \cap \mathcal{O}_1$. Now since $\mathcal{O}_2$ is open and sense, $S_1 \cap \mathcal{O}_2$ is open and nonempty. That is, there exist $r_2 > 0$ and $x_2 \in S_1 \cap \mathcal{O}_2$ such that $S_2 := B(x_2, r_2) \subseteq S_1 \cap \mathcal{O}_2$. Continuing in such fashion, we find $x_n \in S_{n-1} \cap \mathcal{O}_n$ and $r_n \to 0$ such that

$$\overline{S}_{n+1} \subseteq S_n \cap \mathcal{O}_n \subseteq S_1 \cap \mathcal{O}_n.$$

We see that the points $\{x_n\}_{n=1}^\infty$ form a Cauchy sequence, for, if $n, m \geq N$, then $x_n, x_m \in S_N$ and

$$d(x_n, x_m) \leq d(x_n, x_N) + d(x_N, x_m) < 2r_N.$$

Since X is complete, the sequence $\{x_n\}_{n=1}^{\infty}$ converges to some point $x^* \in X$. Since for $n > N$ we have

$$x_n \in \overline{S}_N \subseteq S \cap \mathcal{O}_n,$$

we can let $n \to \infty$ to conclude that $x^* \in \overline{S}_{n+1} \subseteq S \cap \mathcal{O}_n$. Since this is for all $n \in \mathbb{N}$, the set $\bigcap_{n=1}^{\infty} \mathcal{O}_n$ intersects S. ☐

Corollary 1.7.3. *If a complete metric space is expressed as a countable union of closed sets, then one of the closed sets has a nonempty interior.*

Proof. Let X be a complete metric space, and suppose by contradiction that $X = \bigcup_{n=1}^{\infty} F_n$, where each F_n is a closed set having empty interior. The sets $\mathcal{O}_n := X \setminus F_n$ are open and dense. Hence by Baire's Theorem (1.7.2), $\bigcap_{n=1}^{\infty} \mathcal{O}_n$ is open and dense. In particular, it is nonempty. Thus there exists $x \in \bigcap_{n=1}^{\infty} \mathcal{O}_n$, so that $x \in X \setminus \bigcap_{n=1}^{\infty} F_n$, a contradiction. ☐

Definition 1.7.4 (Nowhere Dense). *A subset A in a topological space X is said to be* **nowhere dense** *in X if its closure has empty interior, that is, if $\overline{A}^{\circ} = \emptyset$.*

Example 1.7.5. *Singletons are nowhere dense in $\mathbb{R}$, and the set of irrational points on the horizontal axis in $\mathbb{R}^2$ is nowhere dense in $\mathbb{R}^2$.*

Definition 1.7.6 (Category I). *A set that is a countable union of nowhere dense sets in a space X is said to be of* **category I** *in X.*

Definition 1.7.7 (Category II). *A set that is not of category I in a space X is said to be of* **category II** *in X.*

Example 1.7.8. *Note that the set of irrationals is of category II in $\mathbb{R}$ but the previous example shows that it is of category I in $\mathbb{R}^2$.*

Theorem 1.7.9 (Banach–Steinhaus Theorem). *Let X be a Banach space and Y a normed linear space, and let $\{T_\alpha\}_{\alpha \in A}$ be a family of bounded linear operators in $\mathcal{L}(X, Y)$. Then $\sup_{\alpha \in A} \|T_\alpha\| < \infty$ if and only if the set $\{x \in X : \sup_{\alpha \in A} \|T_\alpha x\| < \infty\}$ is of the second category in X.*

Proof. Define the set $F := \{x \in X : \sup_{\alpha \in A} \|T_\alpha x\| < \infty\}$.

Assume first that $\sup_{\alpha \in A} \|T_\alpha\| < \infty$, and put $M := \sup_{\alpha \in A} \|T_\alpha\|$. Then, for all $x \in X$, we have

$$\|T_\alpha x\| \le \|T_\alpha\| \|x\| \le M\|x\| < \infty,$$

so that $x \in F$ for all $x \in X$. Then $F = X$, and, since X is of the second category in X, this proves one direction of the theorem.

To show the converse, assume that F is of the second category in X and, for each $n \in \mathbb{N}$, define

$$F_n := \{x \in X : \sup_{\alpha \in A} \|T_\alpha x\| \le n\}.$$

Note that $F = \bigcup_{n=1}^{\infty} F_n$. Since F is of the second category, and each F_n is a closed subset of X, we have by corollary (1.7.3) that there exists some $N \in \mathbb{N}$ such that F_N has nonempty interior. That is, there exists $x_0 \in F_N$ and $\epsilon > 0$ such that

$$B[x_0, \epsilon] := \{x \in X : \|x - x_0\| \le \epsilon\} \subseteq F_N.$$

For any $x \in X$ satisfying $\|x\| \le 1$ we have $x_0 + \epsilon x \in B[x_0, \epsilon]$, for

$$\|(x_0 + \epsilon x) - x_0\| = \|\epsilon x\| = \epsilon\|x\| \le \epsilon.$$

Thus for all $\alpha \in A$,

$$\begin{aligned}
\|T_\alpha x\| &= \left\| T_\alpha \left[\frac{x_0 + \epsilon x - x_0}{\epsilon} \right] \right\| = \frac{1}{\epsilon}\|T_\alpha[(x_0 + \epsilon x) - x_0]\| \\
&\le \frac{1}{\epsilon}\|T_\alpha(x_0 + \epsilon x)\| + \frac{1}{\epsilon}\|T_\alpha x_0\| \le \frac{2N}{\epsilon}.
\end{aligned}$$

Taking the supremum, we find for all $\alpha \in A$ that

$$\|T_\alpha\| = \sup_{\|x\| \le 1} \|T_\alpha x\| \le \frac{2N}{\epsilon}.$$

This completes the proof. $\square$

Theorem 1.7.10 (The Principle of Uniform Boundedness)**.** *Let X be a Banach space and Y a normed linear space, and let $\{T_\alpha\}_{\alpha \in A}$ a family of bounded linear operators in $\mathcal{L}(X, Y)$. If $\sup_{\alpha \in A} \|T_\alpha x\| < \infty$ for each $x \in X$, then $\sup_{\alpha \in A} \|T_\alpha\| < \infty$.*

Proof. If $\sup_{\alpha \in A} \|T_\alpha x\| < \infty$ for every $x \in X$, then the set $\{x \in X : \sup_{\alpha \in A} \|T_\alpha x\| < \infty\}$ is all of X, and is thus of the second category in X. The result follows immediately from the Banach–Steinhaus Theorem (1.7.9). $\square$

Example 1.7.11. *Consider the space $\mathcal{C}[0, 1]$. We are going to show that most members of $\mathcal{C}[0, 1]$ are not differentiable at any given point. Select a point ξ in the open interval $(0, 1)$. For small positive values $h > 0$ we define a linear functional $\phi_h : \mathcal{C}[0, 1] \to \mathbb{R}$ by*

$$\phi_h(x) := \frac{x(\xi + h) - x(\xi - h)}{2h}$$

for $x \in \mathcal{C}[0, 1]$. Note ϕ_h is linear:

$$\begin{aligned}
\phi_h(\alpha x + \beta y) &= \frac{(\alpha x + \beta y)(\xi + h) - (\alpha x + \beta y)(x - h)}{2h} \\
&= \frac{(\alpha x)(\xi + h) + (\beta y)(\xi + h) - (\alpha x)(\xi - h) - (\beta y)(\xi - h)}{2h} \\
&= \alpha \frac{x(\xi + h) - x(\xi - h)}{2h} + \beta \frac{y(\xi + h) - y(\xi - h)}{2h} \\
&= \alpha \phi_h(x) + \beta \phi_h(y).
\end{aligned}$$

Also note $\|\phi_h\| = 1/h$:

$$\begin{aligned}
\|\phi_h\| &= \sup_{\|x\| \le 1} \|\phi_h(x)\| = \sup_{\|x\| \le 1} \left\| \frac{x(\xi + h) - x(\xi - h)}{2h} \right\| \le \sup_{\|x\| \le 1} \left(\left\| \frac{x(\xi + h)}{2h} \right\| \left\| \frac{x(\xi - h)}{2h} \right\| \right) \\
&= \frac{1}{2h} + \frac{1}{2h} = \frac{1}{h}.
\end{aligned}$$

Moreover, we have equality and the supremum is attained if we define $x : [0,1] \to \mathbb{R}$ *by*

$$x(t) := \begin{cases} -1, & 0 \le t \le \xi - h, \\ \frac{t}{h} - \frac{\xi}{h}, & \xi - h < t < \xi + h, \\ 1, & \xi + h \le t \le 1. \end{cases}$$

Since $\sup_h \|\phi_h\| = \infty$*, by the Banach–Steinhaus Theorem, the set of all* $x \in \mathcal{C}[0,1]$ *such that* $\sup_h |\phi_h(x)| < \infty$ *is of the first category. Hence the set of* $x \in \mathcal{C}[0,1]$ *for which* $\sup_h |\phi_h(x)| = \infty$ *is of the second category in* $\mathcal{C}[0,1]$*. That is, the set of functions in* $\mathcal{C}[0,1]$ *that are not differentiable at* ξ *is of the second category in* $\mathcal{C}[0,1]$*.*

Theorem 1.7.12. *Let* $\{T_n\}_{n=1}^{\infty}$ *be a sequence of bounded linear operators from a Banach space* X *into a normed linear space* Y*. Then* $\lim_{n\to\infty} T_n x = 0$ *for all* $x \in X$ *if and only if the following two conditions are satisfied:*

(1) $\sup_{n\in\mathbb{N}} \|T_n\| < \infty$;

(2) $\lim_{n\to\infty} T_n u = 0$ *for every* u *in some fundamental subset of* X*.*

Proof. First assume that $\lim_{n\to\infty} T_n x = 0$ for all $x \in X$. Then clearly for any fundamental subset $V \in X$, we have $\lim_{n\to\infty} T_n u =$ for all $u \in V$. Moreover, since each $T_n, n \in \mathbb{N}$ is bounded and $\lim_{n\to\infty} T_n x = 0$ for all $x \in X$ by the assumption, we have by the Principle of Uniform Boundedness (1.7.10), we have $\sup_{n\in\mathbb{N}} \|T_n\| < \infty$.

For the converse, suppose that $\sup_{n\in\mathbb{N}} \|T_n\| \le M < \infty$ and that there exists a fundamental subset V of X such that for all $u \in V$, $\lim_{n\to\infty} T_n u = 0$.

If $x \in \text{span}(V)$, then there exist $u_1, \ldots, u_N \in V$ and $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$ such that

$$x = \sum_{j=1}^{N} \lambda_j u_j.$$

Thus we find

$$\lim_{n\to\infty} T_n x = \lim_{n\to\infty} T_n \left(\sum_{j=1}^{N} \lambda_j u_j \right) = \lim_{n\to\infty} \sum_{j=1}^{N} \lambda_j T_n u_j = 0.$$

Now let $x \notin \text{span}(V)$ and fix $\epsilon > 0$. Since V is fundamental in X, we can choose $u \in \text{span}(V)$ such that

$$\|x - u\| < \frac{\epsilon}{2M}.$$

Moreover, by the assumption that $\lim_{n\to\infty} T_n u = 0$, there exists a positive integer N such that for all $n \ge N$, we have

$$\|T_n u\| < \epsilon/2.$$

Thus for all $n \ge N$, it follows

$$\begin{aligned} \|T_n x\| &= \|T_n x - T_n u + T_n u\| \le \|T_n(x-u)\| + \|T_n u\| \\ &\le \|T_n\|\|x-u\| + \epsilon/2 < M\left(\frac{\epsilon}{2M}\right) + \epsilon/2 \\ &= \epsilon. \end{aligned}$$

Since ϵ was arbitrary, this completes the proof. □

Example 1.7.13. *Recall the limit definition of the definite Riemann integral of a continuous function x defined on the interval $[a, b]$:*

$$\int_a^b x(t)\ dt = \lim_{n\to\infty} \sum_{j=1}^{n} x\left(a + j\frac{b-a}{n}\right)\left[\frac{b-a}{n}\right].$$

Consider the problem of approximating functions $\psi : \mathcal{C}[a, b] \to \mathbb{R}$ that have the form

$$\psi(x) = \int_a^b x(t)w(t)\ dt,$$

where $x \in \mathcal{C}[a, b]$ and in which w is a fixed integrable **weight function.** *We want to approximate ψ by a sequence of functionals ϕ_n having the form*

$$\phi_n(x) = \sum_{j=1}^{n} A_{nj}x(t_{nj}),$$

where $x \in \mathcal{C}[a, b]$. Note that ϕ_n is simply a linear combination of point–evaluation functionals.

We show that

$$\|\phi_n\| = \sum_{j=1}^{n} |A_{nj}|.$$

First observe

$$\|\phi_n\| = \sup_{\|x\|\le 1} \|\phi_n(x)\| = \sup_{\|x\|\le 1} \left\|\sum_{j=1}^{n} A_{nj}x(t_{nj})\right\| \le \sup_{\|x\|\le 1} \sum_{j=1}^{n} |A_{nj}|\|x(t_{nj})\| = \sum_{j=1}^{n} |A_{nj}|.$$

To see the reverse inequality, we assume that for each n, $\{t_{n1}, t_{n2}, \ldots, t_{nn}\}$ is a set of n mutually distinct arguments in $[a, b]$. Let x^* be any continuous function defined on $[a, b]$ that satisfies $\|x^*\| \le 1$ and

$$x^*(t_{nj}) = \operatorname{sgn}(A_{nj}), \quad j = 1, 2, \ldots, n.$$

Then

$$\|\phi_n\| = \sup_{\|x\|\le 1} \|\phi_n(x)\| \ge \|\phi_n(x^*)\| = \left\|\sum_{j=1}^{n} A_{nj}x^*(t_{nj})\right\| = \sum_{j=1}^{n} |A_{nj}|.$$

It follows $\|\phi_n\| = \sum_{j=1}^{n} |A_{nj}|$.

Theorem 1.7.14. *Let $\psi : \mathcal{C}[a, b] \to \mathbb{R}$ be a continuous linear functional defined by*

$$\psi(x) = \int_a^b x(t)w(t)\ dt$$

and let $\phi_n : \mathcal{C}[a, b] \to \mathbb{R}$, $n \in \mathbb{N}$ be a family of continuous linear functionals of the form

$$\phi_n(x) = \sum_{j=1}^{n} A_{nj}x(t_{nj}).$$

Then $\{\phi_n\}_{n=1}^{\infty}$ converges to ψ for each $x \in \mathcal{C}[a, b]$ if and only if the following two conditions are satisfied:

(1) $\sup_{n\in\mathbb{N}} \sum_{j=1}^{n} |A_{nj}| < \infty$;

(2) $\{\phi_n(x)\}_{n=1}^{\infty}$ *converges to* $\psi(x)$ *for all monomial functions* $x(t) = t^k$, $k = 0, 1, 2, \ldots$.

Proof. Consider the sequence of functionals $\{\psi - \phi_n\}_{n=1}^{\infty}$. The norm of ψ is

$$\|\psi\| = \sup_{\|x\|\leq 1} \|\psi(x)\| = \sup_{\|x\|\leq 1} \left| \int_a^b x(t)w(t)\ dt \right| \leq \sup_{\|x\|\leq 1} \int_a^b |x(t)||w(t)|\ dt = \int_a^b |w(t)|\ dt.$$

Consequently condition (1) is equivalent to the condition

$$\sup_{n\in\mathbb{N}} \|\psi - \phi_n\| < \infty.$$

Note by the Weierstrass Approximation Theorem, the functions $h_k : [a,b] \to \mathbb{R}$ defined by $h_k(t) = t^k$, $k = 0, 1, 2, \ldots$, form a fundamental set in $\mathcal{C}[a,b]$. Then the result follows immediately from the preceding theorem (1.7.12). □

1.8. **The Open Mapping and Closed Mapping Theorems.**

Definition 1.8.1 (Closed Function)**.** *Let X and Y be normed linear spaces. A function $f : X \to Y$ is said to be* **closed** *if f is closed as a subset of $X \times Y$, that is, if the set*

$$\{(x, f(x)) : x \in X\}$$

is closed in $X \times Y$.

In terms of sequences, the closed property of f is that if $x_n \to x$ and $f(x_n) \to y$, then $y = f(x)$. Clearly, every continuous map is closed.

An example of a linear transformation that is closed but *not* continuous is the derivative operator D acting on the set of differentiable functions in $\mathcal{C}[a,b]$ and mapping into $\mathcal{C}[a,b]$. The following theorem shows that if $x_n \to x$ and $Dx_n \to y$, then $y = Dx$. We denote by $\mathcal{C}^1[a,b]$ the set of all continuously differentiable functions on $[a,b]$.

Theorem 1.8.2. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathcal{C}^1[a,b]$ and suppose that $\|x_n - x\|_\infty \to 0$ and $\|x_n' - y\|_\infty \to 0$ as $n \to \infty$. Then $y \in \mathcal{C}[a,b]$ and $x' = y$.*

Proof. Since $x_n \in \mathcal{C}^1[a,b]$, we have $x_n' \in \mathcal{C}[a,b]$ for all $n \in \mathbb{N}$. Since $\{x_n'\}_{n=1}^{\infty}$ converges to y with respect to the uniform norm by the assumption, and since $(\mathcal{C}[a,b], \|\cdot\|_\infty)$ is a Banach space (1.2.22), it follows $y \in \mathcal{C}[a,b]$. Moreover, by the fundamental theorem of calculus,

$$\begin{aligned}\int_a^s y(t)\ dt &= \int_a^s \left(\lim_{n\to\infty} x_n'(t)\right) dt = \lim_{n\to\infty} \int_a^s x_n'(t)\ dt = \lim_{n\to\infty} x_n(s) - x_n(a) \\ &= x(s) - x(a).\end{aligned}$$

Differentiation with respect to s now yields $y(t) = x'(t)$. □

Recall from introductory real analysis that the uniform convergence of the derivatives is necessary. In fact, uniform convergence of a sequence of continuously differentiable functions does not guarantee that their derivatives converge even pointwise, for consider

$$x_n(t) = \frac{1}{n}\sin(nt), \quad x_n'(t) = \cos(nt).$$

Definition 1.8.3 (Open Mapping)**.** *Let X and Y be normed linear spaces. A function $f : X \to Y$ is said to be* **open** *if f maps open sets in X to open sets in Y, that is, if for every open set $U \in X$, $f(U)$ is open in Y.*

Theorem 1.8.4 (The Open Mapping Theorem)**.** *If a closed linear transformation maps one Banach space onto another, then it is an open map.*

Proof. Let X and Y be Banach spaces and let $L : X \to Y$ be a surjective closed linear map. We denote by $S_X := B_X(0,1) = \{x \in X : \|x\| < 1\}$ the open unit ball in X and $S_Y := B_Y(0,1) = \{y \in Y : \|y\| < 1\}$ the open unit ball in Y.

We first show that $S_Y \subseteq \overline{L(tS_x)}$ for some $t > 0$. Since L is surjective by hypothesis,

$$Y = L(X) = L\left(\bigcup_{n=1}^{\infty} B_X(0,n)\right) = \bigcup_{n=1}^{\infty} L(B_X(0,n)) = \bigcup_{n=1}^{\infty} L(nS_X).$$

Since Y is complete, it follows by Baire's Theorem (1.7.2) that one of the sets $\overline{L(nS_X)} = \overline{L(B_X(0,n))}$ has nonempty interior, say $\overline{L(mS_X)} = \overline{L(B_X(0,m)}$ has nonempty interior. By definition, then, there exists $v \in \overline{L(mS_X)}$ and $r > 0$ such that

$$B_Y(v,r) = \{y \in Y : \|y - v\| < r\} = v + rS_Y \subseteq \overline{L(mS_X)}.$$

We see that $v \in \overline{L(mS_X)}$, and thus if $\|y\| < 1$, we have

$$\|ry\| = \|ry + v - v\| \leq \|ry + v\| + \|v\| < r + r = 2r,$$

so that $ry \in B(v, 2r) \subseteq \overline{L(2mS_x)}$. It follows that for all $y \in S_Y$, $y \in \overline{L(tS_X)}$ for $t := 2m/r$. This shows $S_Y \subseteq \overline{L(tS_X)}$.

We now show that $S_Y \subseteq L(2tS_X)$. Fix $y \in S_Y$. Choose a sequence of positive numbers δ_n such that $\sum_{n=1}^{\infty} \delta_n < 1$. Since $y \in \overline{L(tS_X)}$, it follows by the closure of L that there exists $x_1 \in tS_X$ such that $\|y - Lx_1\| < \delta_1$. Since

$$y - Lx_1 \in B_Y(0, \delta_1) = \delta_1 S_Y \subseteq \overline{L(\delta_1 t S_X)},$$

there is a point $x_2 \in \delta_1 t S_X$ such that $\|y - Lx_1 - Lx_2\| < \delta_2$. We continue in this fashion, obtaining a sequence $x_1, x_2, \ldots$ such that

$$x_n \in \delta_{n-1} t S_X$$

and whose partial sums $S_n := x_1 + \cdots + x_n$ have the property

$$\|y - LS_n\| = \|y - L(x_1 + \cdots + x_n)\| = \|y - Lx_1 - Lx_2 + \cdots - Lx_n\| < \delta_n.$$

Also, we have for the sequence $\{S_n\}_{n=1}^{\infty}$ of partial sums that

$$\begin{aligned}
\|S_n\| &= \|x_1 + \cdots + x_n\| \leq \|x_1\| + \cdots + \|x_n\| \\
&\leq t + \delta_1 t + \delta_2 t + \cdots + \delta_{n-1} t \\
&= t\left(1 + \sum_{j=1}^{n-1} \delta_{j=1}\right) \\
&< 2t.
\end{aligned}$$

Moreover, the sequence $\{S_n\}_{n=1}^{\infty}$ is Cauchy, for we see that

$$\begin{aligned}
\|S_{n+k} - S_n\| &= \left\|\sum_{j=1}^{n+k} x_j - \sum_{j=1}^{n} x_j\right\| = \|x_{n+1} + x_{n+2} + \cdots + x_{n+k}\| \\
&\leq \|x_{n+1}\| + \|x_{n+2}\| + \cdots + \|x_{n+k}\| \\
&\leq t\delta_n + t\delta_{n+1} + \cdots + t\delta_{n+k-1}
\end{aligned}$$

$$< t \sum_{j=n}^{\infty} \delta_j.$$

Since X is complete, $\{S_n\}_{n=1}^{\infty}$ converges to s for some $s \in X$. Also, since L is closed and $LS_n \to y$, we have $y = Ls$. Furthermore, since $s = \sum_{j=1}^{\infty} x_j$, we have

$$\|s\| = t\left(1 + \sum_{j=1}^{\infty} \delta_j\right) < 2t.$$

Hence $s \in 2tS_X$, so that $y = Ls \in L(2tS_X)$. Since $y \in S_Y$ was arbitrary, it follows $S_Y \subseteq L(2tS_X)$.

Lastly, to complete the proof, we show that $L(U)$ is open in Y whenever U is open in X. It suffices to show that for any $y \in L(U)$, there exists $\epsilon > 0$ such that $y + \epsilon S_Y \subseteq L(U)$. Let $y \in L(U)$ be arbitrary. Then there exists $x \in U$ such that $y = Lx$. Since U is open, there exists $\delta > 0$ such that

$$B(x, \delta) = x + \delta S_X \subseteq U.$$

Then

$$L(x + \delta S_X) = Lx + \delta L(S_X) = y + \delta L(S_X) \subseteq L(U).$$

By the preceding arguments, we have that $S_X \subseteq L(2tS_X)$. Thus $\frac{\delta}{2t} S_X \subseteq \delta L(S_X)$ and

$$B\left(y, \frac{\delta}{2t}\right) = y + \frac{\delta}{2t} S_X \subseteq L(U).$$

Thus $L(U)$ contains a neighborhood of y, and, since y was arbitrary, $L(U)$ is open.

This completes the proof. □

Corollary 1.8.5. *If an algebraic isomorphism of one Banach space onto another is continuous, then its inverse is continuous.*

Proof. Let X and Y be Banach spaces and let $L : X \to Y$ be a continuous isomorphism. Since L is continuous, L is closed. By the open mapping theorem (1.8.4), L is an open map.

Note that since L is bijective by the assumption, L has a well–defined inverse $L^{-1} : Y \to X$, $L^{-1} : y = L(x) \mapsto x$. Let U be open in X. Then, since L is open, $(L^{-1})^{-1}(U) = L(U)$ is open in Y, which completes the proof. □

Corollary 1.8.6. *Let X be a linear space and suppose that $\|\cdot\|_a, \|\cdot\|_b$ are two norms on X such that $(X, \|\cdot\|_a)$ and $(X, \|\cdot\|_b)$ are both Banach spaces and*

$$\|x\|_a \leq \|x\|_b$$

for all $x \in X$. Then $\|\cdot\|_a$ and $\|\cdot\|_b$ are equivalent, in the sense that there exists $\lambda > 0$ such that

$$\lambda\|x\|_b \leq \|x\|_a \leq \|x\|_b$$

for all $x \in X$.

Proof. First note that $\|x\|_a \leq \|x\|_b$ follows by the assumption.

Denote by I the identity map $I : (X, \|\cdot\|_b) \to (X, \|\cdot\|_a)$ defined by $Ix = x$ for all $x \in X$. Then, for all $x \in X$, we have

$$\|Ix\|_a = \|x\|_a \leq \|x\|_b.$$

Consequently, I is bounded and thus continuous. By the preceding corollary (1.8.5), the inverse $I^{-1} : (X, \|\cdot\|_a) \to (X, \|\cdot\|_b)$ is continuous, and hence, bounded. Thus for all $x \in X$ there exists $\lambda > 0$ such that

$$\|x\|_b = \|I^{-1}x\|_b \leq \frac{1}{\lambda}\|x\|_a.$$

Multiplying by λ, we find

$$\lambda\|x\|_b \leq \|x\|_a \leq \|x\|_b$$

for all $x \in X$, which completes the proof. □

Theorem 1.8.7 (The Closed Graph Theorem)**.** *Let X and Y be Banach spaces. If a linear map $L : X \to Y$ is closed, then L is continuous.*

Proof. Let $L : X \to Y$ be a closed linear operator. Define a new norm N on X by

$$N(x) := \|x\|_X + \|Lx\|_Y$$

for all $x \in X$. To see that N is in fact a norm, note that N is nonnegative and $N(x) = 0$ if and only if $x = 0$. Second, for all $\alpha \in \mathbb{R}$ and $x \in X$,

$$N(\alpha x) = \|\alpha x\|_X + \|L(\alpha x)\|_Y = |\alpha|\|x\|_X + |\alpha|\|Lx\|_Y = |\alpha|(\|x\|_X + \|Lx\|_Y) = |\alpha|N(x).$$

The triangle inequality also follows:

$$\begin{aligned} N(x+y) = \|x+y\|_X + \|L(x+y)\|_Y &\leq \|x\|_X + \|y\|_Y + \|Lx + Ly\|_Y \\ &\leq (\|x\|_X + \|Lx\|_Y) + (\|y\| + \|Ly\|_Y) = N(x) + N(y). \end{aligned}$$

Consequently N is a norm.

We show that (X, N) is complete. Let $\{x_n\}_{n=1}^\infty$ be Cauchy in (X, N). Since

$$\|x\|_X \leq N(x) \quad \text{and} \quad \|Lx\|_Y \leq N(x)$$

for all $x \in X$, $\{x_n\}_{n=1}^\infty$ is Cauchy in $(X, \|\cdot\|_X)$ and $\{Lx_n\}_{n=1}^\infty$ is Cauchy in $(Y, \|\cdot\|)$. Since X and Y are complete, $\{x_n\}_{n=1}^\infty$ converges to some $x \in X$ and $\{Lx_n\}_{n=1}^\infty$ converges to some $y \in Y$. Moreover, since L is closed, $y = Lx$. Then

$$N(x_n - x) = \|x_n - x\|_X + \|L(x_n - x)\|_Y = \|x_n - x\| + \|Lx_n - y\|_Y \to 0$$

as $n \to \infty$. Hence (X, N) is complete.

Finally, by the preceding corollary (1.8.6), there exists $\lambda > 0$ such that $\|N(x)\| \leq \lambda\|x\|$ for all $x \in X$. It follows

$$\|Lx\|_Y \leq \|x\|_X + \|Lx\|_Y = N(x) \leq \lambda\|x\|_X$$

for all $x \in X$. Hence L is bounded, and thus continuous, which completes the proof. □

Example 1.8.8. *Note that the derivative operator $D : (\mathcal{C}^1[a,b], \|\cdot\|_\infty) \to (\mathcal{C}[a,b], \|\cdot\|_\infty)$ was shown to be a closed unbounded operator in a preceding example. Since $\mathcal{C}[a,b]$ is a Banach space under the infinity norm, we can conclude that the space $\mathcal{C}^1[a,b]$ does not form a Banach space under the infinity norm.*

Lemma 1.8.9. *A normed linear space X is a Banach space if and only if every absolutely convergent series is convergent.*

Proof. Let X be a normed linear space.

First assume that X is a Banach space, and let $\sum_{n=1}^{\infty} x_n$ converge absolutely. Denote by $\{S_n\}_{n=1}^{\infty}$ the sequence of partial sums

$$S_n := x_1 + x_2 + \cdots + x_n, \quad n = 1, 2, \ldots.$$

Since $\sum_{n=1}^{\infty} \|x_n\| < \infty$ by the assumption, for any $\epsilon > 0$ there exists a positive integer N such that $\sum_{n=N}^{\infty} \|x_n\| < \epsilon$. Consequently $\{S_n\}_{n=1}^{\infty}$ is Cauchy, for we see that, for all $n, n = k \geq N, k \geq 1$,

$$\|S_{n+k} - S_n\| = \left\| \sum_{j=1}^{n+k} x_j - \sum_{j=1}^{n} x_j \right\| = \left\| \sum_{j=n+1}^{n+k} x_j \right\| \leq \sum_{j=n+1}^{n+k} \|x_j\| < \epsilon.$$

Since X is complete, $\{S_n\}_{n=1}^{\infty}$ converges to some $S \in X$, which completes the proof for this direction.

Conversely, suppose that every absolutely convergent series in X converges and let $\{x_n\}_{n=1}^{\infty}$ be Cauchy in X. Then for every $k \in \mathbb{N}$, there exists a positive integer N_k such that for all $n, m \geq N_k$, we have

$$\|x_n - x_m\| < \frac{1}{k^2}.$$

Without loss of generality, we may assume that the sequence $\{N_k\}_{k=1}^{\infty}$ is increasing. Then the series $\sum_{k=1}^{\infty} x_{N_{k+1}} - x_{N_k}$ converges absolutely, for we see that

$$\sum_{k=1}^{\infty} \|x_{N_{k+1}} - x_{N_k}\| < \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty,$$

since the RHS is a convergent $p-$series. Evidently the series $\sum_{k=1}^{\infty} x_{N_{k+1}} - x_{N_k}$ converges by the assumption. Moreover, we see that for each $k \in \mathbb{N}$,

$$x_{N_k} = x_{N_1} + (x_{N_2} - x_{N_1}) + \cdots + (x_{N_k} - x_{N_{k-1}}) = x_{N_1} + \sum_{j=1}^{k} x_{j+1} - x_j.$$

Thus the sequence $\{x_{N_k}\}_{k=1}^{\infty}$ converges to some $x \in X$. Observe

$$\|x_n - x\| \leq \|x_n - x_{N_k}\| + \|x_{N_k} - x\| \to 0$$

as $n \to \infty$ for some N_k large enough. This completes the proof. □

Theorem 1.8.10. *A normed linear space that is the image of a Banach space under a bounded linear open map is also a Banach space.*

Proof. Let X be a Banach space and let $L : X \to Y$ be a bounded linear open map onto Y. By the lemma (1.8.9), it suffices to prove that every absolutely convergent series in Y is convergent. Let $\{y_n\}_{n=1}^{\infty}$ be a sequence in Y such that $\sum_{n=1}^{\infty} \|y_n\| < \infty$. Since L is open, there exists $x_n \in X$ such that $Lx_n = y_n$ and $\|x_n\| \leq C\|y_n\|$ for some $C > 0$. Thus

$$\sum_{n=1}^{\infty} \|x_n\| \leq C \sum_{n=1}^{\infty} \|y_n\| < \infty.$$

Since X is complete, it follows by the lemma (1.8.9) that the series $\sum_{n=1}^{\infty} x_n$ converges. Finally, since L is continuous and linear,

$$\sum_{n=1}^{\infty} y_n = \sum_{n=1}^{\infty} Lx_n = L\left(\sum_{n=1}^{\infty} x_n\right)$$

converges. □

Definition 1.8.11 (Adjoint Operator)**.** *Let X and Y be normed linear spaces and let $L : X \to Y$ be a bounded linear operator. The* **adjoint operator** *of L is the map $L^* : Y^* \to X^*$ defined by $L^*\phi = \phi \circ L$ for all bounded linear functionals $\phi \in Y^*$.*

Proposition 1.8.12. *Let $L : X \to Y$ be a bounded linear operator. Then $L^* : Y^* \to X^*$ is also a bounded linear operator.*

Proof. We first show that L^* is linear. Let $\phi, \psi \in Y^*$ and let $\alpha, \beta \in \mathbb{R}$. Then, by the linearity of ϕ and ψ,

$$\begin{aligned} L^*(\alpha\phi + \beta\psi) &= (\alpha\phi + \beta\psi) \circ L = (\alpha\phi) \circ L + (\beta\psi) \circ L = \alpha(\phi \circ L) + \beta(\psi \circ L) \\ &= \alpha L^*\phi + \beta L^*\psi. \end{aligned}$$

To see that L^* is bounded, recall that for any $T \in \mathcal{L}(\mathcal{L}(Y, \mathbb{R}), \mathcal{L}(X, \mathbb{R}))$, we have

$$\|T\| = \sup_{\|\phi\|_{\mathcal{L}(Y,\mathbb{R})} \le 1} \|T\phi\|_{\mathcal{L}(X,\mathbb{R})} = \sup_{\|\phi\|_{\mathcal{L}(Y,\mathbb{R})} = 1} \|T\phi\|_{\mathcal{L}(X,\mathbb{R})}.$$

We see that

$$\|L^*\| = \sup_{\|\phi\|=1} \|L^*\phi\| = \sup_{\|\phi\|=1} \sup_{\|x\|\le 1} |(L^*\phi)(x)| = \sup_{\|x\|\le 1} \sup_{\|\phi\|=1} |\phi(Lx)|.$$

Since $\phi \in Y^*$ with $\|\phi\| = 1$, it follows by Corollary (1.6.25) that $\sup_{\|\phi\|=1} |\phi(Lx)| = \|Lx\|$. Hence,

$$\|L^*\| = \sup_{\|x\|\le 1} \|Lx\| = \|L\|.$$

This shows that L^* is bounded, and, in particular, L^* has the same norm as L in operator norm. □

In a finite–dimensional space, an operator L can be represented by a matrix A. This requires the prior selection of bases for the domain and range of L. The adjoint operator L^* is represented by the complex conjugate matrix A^H.

Definition 1.8.13 (Range Space)**.** *Let X and Y be normed linear spaces and let $L : X \to Y$ be a linear operator. The* **range space** *of L, denoted by $\mathcal{R}(L)$, is the set*

$$\mathcal{R}(L) := L(X) = \{y \in Y : y = Lx\}.$$

Definition 1.8.14 (Null Space)**.** *Let X and Y normed linear spaces and let $L : X \to Y$ be a linear operator. The* **null space** *of L, denoted by $\mathcal{N}(L)$, is the set*

$$\mathcal{N}(L) := \ker(L) = \{x \in X : Lx = 0\}.$$

Theorem 1.8.15. *Let X and Y be normed linear spaces and let $L \in \mathcal{L}(X, Y)$. Then $\mathcal{R}(L)$ is dense if and only if L^* is injective.*

Proof. Let $L : X \to Y$ be continuous.

Suppose that $\mathcal{R}(L)$ is dense in Y. Then $L(X)$ is a fundamental subset of Y, so it follows by Theorem (1.6.29) that $L(X)^{\perp} = \{0\}$. That is, if $\phi \in L(X)^{\perp}$, then $\phi \equiv 0$. Thus, if $\phi(Lx) = 0$ for all $x \in X$, then $\phi \equiv 0$. But then $L^*(\phi) = \phi(Lx) = 0$ for all $\phi \in L(X)^{\perp}$, so that $\ker(L^*) = \{0\}$. This shows that L^* is injective.

For the converse, assume that L^* is injective. Then $\ker(L^*) = \{0\}$, so that for all $\phi \in L(X)^{\perp}$, we have $L^*(\phi) = \phi(Lx) = 0$ if and only if $\phi \equiv 0$. Thus $L(X)^{\perp} = \{0\}$, which implies by Theorem (1.6.29) that $L(X)$ is dense in Y. ☐

Remark. *Let X be a normed linear space and let $U \subseteq X^*$. We use the notation*

$$U_{\perp} := \{x \in X : \phi(x) = 0 \text{ for all } \phi \in U\}.$$

Theorem 1.8.16 (Closed Range Theorem)**.** *Let X and Y be normed linear spaces and let $L \in \mathcal{L}(X, Y)$. Then $\overline{\mathcal{R}(L)} = \mathcal{N}(L^*)_{\perp}$.*

Proof. We first show $\overline{\mathcal{R}(L)} \subseteq \mathcal{N}(L^*)_{\perp}$. Let $y \in \overline{\mathcal{R}(L)}$. Then there exists a sequence $\{y_n\}_{n=1}^{\infty}$ in $\mathcal{R}(L)$ converging to y. Since each $y_n \in \mathcal{R}(L)$ for every $n \in \mathbb{N}$, there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in X such that $y_n = Lx_n$ for each $n \in \mathbb{N}$. To show that $y \in \mathcal{N}(L^*)_{\perp}$, we show that $\phi(y) = 0$ for any $\phi \in \mathcal{N}(L^*)$. Given $\phi \in \mathcal{N}(L^*)$, we have

$$\begin{aligned}\phi(y) &= \phi\left(\lim_{n\to\infty} y_n\right) = \lim_{n\to\infty} \phi(y_n) = \lim_{n\to\infty} \phi(Lx_n)\\ &= \lim_{n\to\infty} (\phi \circ L)(x_n) = \lim_{n\to\infty} L^*(\phi(x_n)) = 0,\end{aligned}$$

which shows $\overline{\mathcal{R}(L)} \subseteq \mathcal{N}(L^*)_{\perp}$.

We now show $\mathcal{N}(L^*)_{\perp} \subseteq \overline{\mathcal{R}(L)}$. We proceed by contrapositive and suppose that $y \notin \overline{\mathcal{R}(L)}$. Since $\overline{\mathcal{R}(L)}$ is a subspace of Y, it follows from Corollary (1.6.18) that there exists a continuous linear functional $\phi \in X^*$ such that $\phi(y) = 1$ and $\phi(z) = 0$ for all $z \in \overline{\mathcal{R}(L)}$. Thus, for all $x \in X$, we have

$$L^*(\phi(x)) = (\phi \circ L)(x) = \phi(Lx) = 0.$$

Consequently $\phi \in \mathcal{N}(L^*)_{\perp}$. But since $\phi(y) \neq 0$, we conclude that $y \notin \mathcal{N}(L^*)_{\perp}$. ☐

Theorem 1.8.17. *Let X and Y be Banach spaces and let $L \in \mathcal{L}(X, Y)$. Then the range of L is closed if and only if L is bounded below, specifically,* $\inf_{\|x\|=1} \|Lx\|_Y > 0$.

Proof. First assume that $\overline{\mathcal{R}(L)}$ is closed. Then $\mathcal{R}(L)$ is a Banach space. Since $L : X \to \mathcal{R}(L)$ is bijective, it follows by Corollary (1.8.5) that L has a continuous inverse $L^{-1} : \mathcal{R}(L) \to X$. Let $y \in \mathcal{R}(L)$. Then there exists $x \in X$ such that $Lx = y$. Since L^{-1} is bounded, we have $\|L^{-1}y\| \leq \|L^{-1}\|\|y\|$, from which it follows

$$\|x\| \leq \|L^{-1}\|\|Lx\|.$$

Hence,

$$\inf_{\|x\|=1} \|Lx\| \geq \frac{1}{\|L^{-1}\|} > 0,$$

which shows that L is bounded below.

For the converse, suppose that L is bounded below, that is, that there exists $c > 0$ such that $\|Lx\| \geq c > 0$ for all $x \in X$ such that $\|x\| = 1$. By homogeneity, $\|Lx\| = c\|x\|$ for all

$x \in X$. To show that $\mathcal{R}(L)$ is closed, let $\{y_n\}_{n=1}^\infty$ be a sequence in $\mathcal{R}(L)$ converging to $y \in Y$. By definition of $\mathcal{R}(L)$, there exists a sequence $\{x_n\}_{n=1}^\infty$ in X such that $Lx_n = y_n$ for each $n \in \mathbb{N}$. Observe for all $n, m \in \mathbb{N}$ that

$$c\|x_n - x_m\| \leq \|L(x_n - x_m)\| = \|Lx_n - Lx_m\| = \|y_n - y_m\|,$$

so that $\{x_n\}_{n=1}^\infty$ is Cauchy by the convergence of $\{y_n\}_{n=1}^\infty$. Since X is complete, $\{x_n\}_{n=1}^\infty$ converges to some point $x \in X$. By continuity,

$$Lx = L\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} L(x_n) = \lim_{n\to\infty} y_n = y,$$

which shows $y \in \mathcal{R}(L)$.

This completes the proof. □

1.9. **Weak Convergence.**

Definition 1.9.1 (Weak Convergence)**.** *A sequence $\{x_n\}_{n=1}^\infty$ in a normed linear space X is said to* **converge weakly** *to a point $x \in X$ if for all $\phi \in X^*$, the sequence $\{\phi(x_n)\}_{n=1}^\infty$ converges to $\phi(x)$.*

If a sequence $\{x_n\}_{n=1}^\infty$ converges weakly to x, we write $x_n \rightharpoonup x$. The usual convergence is typically called *strong* or *norm convergence.*

Proposition 1.9.2. *If a sequence in any normed linear space converges in norm, then it converges weakly.*

Proof. Suppose that $\{x_n\}_{n=1}^\infty$ is a sequence in a normed linear space X converging to $x \in X$. Then, since every $\phi \in X^*$ is continuous, it follows

$$\lim_{n\to\infty} \phi(x_n) = \phi\left(\lim_{n\to\infty} x_n\right) = \phi(x).$$

It follows that $\{x_n\}_{n=1}^\infty$ converges in a weak sense. □

Weak convergence does not imply strong convergence in general.

Example 1.9.3. *Consider the space c_0 of all sequences converging to zero and the standard unit vectors $e_{n_k} := \delta_{nk}$. Recall, since c_0^* and ℓ_1 are algebraically isomorphic, every continuous linear functional on c_0 is of the form*

$$\phi(x) = \sum_{i=1}^\infty \alpha_i x_i$$

for some $\alpha \in \ell_1$. Thus

$$\phi\left(\lim_{n\to\infty} e_n\right) = \lim_{n\to\infty} \phi(e_n) = \lim_{n\to\infty} \alpha_n = 0,$$

since $\{\alpha_n\}_{n=1}^\infty$ converges. On the other hand, the sequence $\{e_n\}_{n=1}^\infty$ does not converge, for

$$\|x_n - x_m\| = 1$$

whenever $n \neq m$.

Lemma 1.9.4. *Any weakly convergent sequence is bounded.*

Proof. Let X be a normed linear space and suppose that $\{x_n\}_{n=1}^{\infty}$ converges in a weak sense to some point $x \in X$. Define a sequence of functionals $\{\hat{x}_n\}_{n=1}^{\infty}$ on X^* by putting

$$\hat{x}_n(\phi) := \phi(x_n)$$

for each $\phi \in X^*$. By the assumption, the sequence $\{\phi(x_n)\}_{n=1}^{\infty}$ converges in $\mathbb{R}$ for each ϕ, and so the sequence $\{\hat{x}_n(\phi)\}_{n=1}^{\infty}$ converges for each $\phi \in X^*$ and is thus bounded. Since X^* is complete and $\sup_{n\in\mathbb{N}} |\hat{x}_n(\phi)| < \infty$, it follows by the Uniform Boundedness Theorem (1.7.10) that there exists $M > 0$ such that $\|\hat{x}_n\| \leq M$ for all $n \in \mathbb{N}$. That is,

$$\sup_{n\in\mathbb{N}} \{|\hat{x}_n(\phi)| : \phi \in X^*,\ \|\phi\| \leq 1\} \leq M.$$

Finally, by Corollary (1.6.25), it follows

$$\|x_n\| \leq M$$

for all $n \in \mathbb{N}$, which completes the proof. □

Theorem 1.9.5. *In any finite–dimensional normed linear space, a sequence converges weakly if and only if it converges strongly.*

Proof. It suffices to show that weak convergence implies strong convergence.

Let X be a normed linear space such that $\dim(X) = n$. Choose a basis $\{b_1, \dots, b_n\}$ for X and let $\phi_1, \dots, \phi_n$ be linear functionals such that for each $x \in X$,

$$x = \sum_{k=1}^{n} \phi_k(x) b_k.$$

Since X is finite–dimensional, we have by (1.5.17) that each ϕ_k, $k = 1, \dots, n$ is continuous.

Now suppose that a sequence $\{x_m\}_{m=1}^{\infty}$ converges weakly to $x \in X$. Fix $\epsilon > 0$ and put $M := \max\{\|b_1\|, \dots, \|b_n\|\}$. Since $\{\phi_k(x_m)\}_{m=1}^{\infty}$ converges to $\phi_k(x)$ for each $k = 1, \dots, n$ by the assumption, there exists a positive integer N such that for all $m \geq N$,

$$|\phi_k(x_m) - \phi_k(x)| < \frac{\epsilon}{nM}$$

for each $k = 1, \dots, n$. Observe that for all $m \geq N$,

$$\begin{aligned} \|x_m - x\| &= \left\|\sum_{k=1}^{n} (\phi_k(x_m) - \phi_k(x)) b_k\right\| = \sum_{k=1}^{n} |\phi_k(x_m) - \phi_k(x)| \|b_k\| < \sum_{k=1}^{n} \left(\frac{\epsilon}{nM}\right) M \\ &= \epsilon. \end{aligned}$$

Hence, $\{x_n\}_{n=1}^{\infty}$ converges to x in norm. □

Theorem 1.9.6. *If a sequence $\{x_n\}_{n=1}^{\infty}$ in a normed linear space X converges in a weak sense to $x \in X$, then a sequence of linear combinations of the terms x_n converges strongly to x.*

Proof. We show that $x \in \overline{\operatorname{span}\{x_1, x_2, \dots\}} =: Y$.

By contradiction, suppose that $x \notin Y$. Since Y is a subspace of X, it follows by Corollary (1.6.18) that there exists a continuous linear functional $\phi \in X^*$ such that $\phi(y) = 0$ for all

$y \in Y$ and $\phi(x) = 1$. On the other hand, we have by the assumption and continuity of ϕ that

$$\phi\left(\lim_{n\to\infty} x_n\right) = \lim_{n\to\infty} \phi(x_n) = 0,$$

a contradiction. □

In fact, a sequence of *convex* linear combinations of $\{x_1, x_2, \dots\}$ converges strongly to x, which can be proved with a separation axiom from point–set topology.

Theorem 1.9.7. *Let X be a normed linear space, and suppose that the sequence $\{x_n\}_{n=1}^\infty$ in X is bounded and that $\{\phi(x_n)\}_{n=1}^\infty$ converges to $\phi(x)$ for all ϕ in a fundamental subset of X^*. Then x_n converges weakly to x.*

Proof. Let $\mathcal{F} \subset X^*$ be a fundamental subset of X^* and let $\psi \in X^*$ be arbitrary.

Fix $\epsilon > 0$. By the assumption, there exists $M > 0$ such that $\|x_n\| \le M$ for all $n \in \mathbb{N}$. Since $\mathcal{F}$ is fundamental in X^*, there exist $\phi_1, \dots, \phi_m \in \mathcal{F}$ and $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ such that

$$\left\|\psi - \sum_{k=1}^m \lambda_k \phi_k\right\| < \frac{\epsilon}{3M}.$$

Put $\phi := \sum_{k=1}^m \lambda_k \phi_k$. Since ϕ is a linear combination of functionals in $\mathcal{F}$, clearly $\{\phi(x_n)\}_{n=1}^\infty$ converges to $\phi(x)$. Thus there exists a positive integer N such that for all $n \ge N$, we have $|\phi(x_n) - \phi(x)| < \epsilon/3$. Then, for all $n \ge N$, observe

$$\begin{aligned}
|\psi(x_n) - \psi(x)| &\le |\psi(x_n) - \phi(x_n)| + |\phi(x_n) - \phi(x)| + |\phi(x) - \psi(x)| \\
&< \|\psi - \phi\|\|x_n\| + \epsilon/3 + \|\psi - \phi\|\|x\| \\
&< \left(\frac{\epsilon}{3M}\right) M + \epsilon/3 + \left(\frac{\epsilon}{3M}\right) \\
&= \epsilon.
\end{aligned}$$

This completes the proof. □

Example 1.9.8. *Fix a real number p in the range $1 \le p < \infty$. The space ℓ_p is defined to be the set of all real sequences $\{x_n\}_{n=1}^\infty$ for which $\sum_{n=1}^\infty |x_n|^p < \infty$. We define a norm on the linear space ℓ_p by the equation*

$$\|x\|_p := \left\{\sum_{n=1}^\infty |x_n|^p\right\}^{1/p}.$$

For $p = \infty$, we take ℓ_∞ to be the space of all bounded sequences, with norm $\|x\|_\infty := \sup_{n\in\mathbb{N}} |x_n|$.

Theorem 1.9.9 (Hölder Inequality, ℓ_p)**.** *Let $1 < p < \infty$, let $1/p + 1/q = 1$, and let $x \in \ell_p$, $y \in \ell_q$. Then*

$$\sum_{n=1}^\infty x_n y_n \le \|x\|_p \|y\|_q.$$

Theorem 1.9.10 (Minkowski Inequality). *If $x, y \in \ell_p$, then*
$$\|x+y\|_p \leq \|x\|_p + \|y\|_q.$$

Proof. If $p = 1$, then
$$\|x+y\|_1 = \sum_{n=1}^{\infty} |x_n + y_n| \leq \sum_{n=1}^{\infty} |x_n| + |y_n| = \sum_{n=1}^{\infty} |x_n| + \sum_{n=1}^{\infty} |y_n| = \|x\|_1 + \|y\|_1.$$

Now assume that $1 < p < \infty$. Then
$$\begin{aligned}\sum_{n=1}^{\infty} |x_n + y_n|^p &\leq \sum_{n=1}^{\infty} \{|x_n| + |y_n|\}^p \leq \sum_{n=1}^{\infty} \{2 \max\{|x_n|, |y_n|\}\}^p \\ &= \sum_{n=1}^{\infty} 2^p \max\{|x_n|^p, |y_n|^p\} \leq 2^p \{|x_n|^p + |y_n|^p\} \\ &< \infty.\end{aligned}$$

Taking the $p-$th root of both sides, this shows $x + y \in \ell_p$.

Now let $1/p + 1/q = 1$. Then $p + q = pq$, and we see that if $x \in \ell_p$, then $|x|^{p-1} \in \ell_q$, because
$$\sum_{n=1}^{\infty} \left\{|x_n|^{p-1}\right\}^q = \sum_{n=1}^{\infty} |x_n|^p < \infty.$$

Thus, by the Hölder inequality,
$$\begin{aligned}\|x+y\|_p^p = \sum_{n=1}^{\infty} |x_n + y_n|^p &= \sum_{n=1}^{\infty} |x_n + y_n|^{p-1} |x_n + y_n| \\ &\leq \sum_{n=1}^{\infty} |x_n + y_n|^{p-1} |x_n| + \sum_{n=1}^{\infty} |x_n + y_n|^{p-1} |y_n| \\ &\leq \left\| |x+y|^{p-1} \right\|_q \|x\|_p + \left\| |x+y|^{p-1} \right\|_q \|y\|_p \\ &= \left\| |x+y|^{p-1} \right\|_q \{\|x\|_p + \|y\|_p\} \\ &= \|x+y\|_q^{p/q} \{\|x\|_p + \|y\|_p\}.\end{aligned}$$

Finally, since $p - p/q = 1$, dividing both sides by $\|x+y\|_p^{p/q}$ gives
$$\|x+y\|_p \leq \|x\|_p + \|y\|_p,$$
which completes the proof. □

Theorem 1.9.11. *The adjoint ℓ_p^* is isometrically isomorphic to ℓ_q, where $1/p + 1/q = 1$, $1 \leq p < \infty$. That is, for each element $\phi \in \ell_p^*$ there exists a unique element $y \in \ell_q$ such that $\phi(x) = \sum_{n=1}^{\infty} x_n y_n$.*

Theorem 1.9.12. *Let x and $\{x^{(n)}\}_{n=1}^{\infty}$ be in ℓ_p. Then $\{x^{(n)}\}_{n=1}^{\infty}$ converges weakly to x if and only if $\|x^{(n)}\|_p$ is bounded and $\{x_k^{(n)}\}_{n=1}^{\infty}$ converges to x_k for each $k \in \mathbb{N}$.*

Proof. First, suppose that $\{x^{(n)}\}_{n=1}^{\infty}$ converges weakly to x. Since weakly convergent sequences are bounded (1.9.4), it follows $\|x^{(n)}\| < \infty$ for all $n \in \mathbb{N}$. To show that $\{x_k^{(n)}\}_{n=1}^{\infty}$

converges to x_k for each fixed $k \in \mathbb{N}$, note by (1.9.11) that every continuous linear functional on ℓ_p has the form

$$\phi(x) := \sum_{n=1}^{\infty} \alpha_n x_n$$

for $\alpha \in \ell_q$, where $1/p + 1/q = 1$. Define a sequence of functionals $\{\phi_m\}_{m=1}^{\infty}$ on ℓ_p by

$$\phi_m(x) := \sum_{k=1}^{\infty} e_k^{(m)} x_k,$$

where $e^{(k)}$ denote the standard basis vectors in ℓ_q. We observe

$$\phi_m(x^{(n)}) = \sum_{k=1}^{\infty} e_k^{(m)} x_k^{(n)} = x_m^{(n)}.$$

But since $\phi_m(x^{(n)}) \to x_m$ by the assumption, so does $x_m^{(n)}$. This completes the proof for this direction.

Next, suppose that $\|x^{(n)}\|_p$ is bounded and $\{x_k^{(n)}\}_{n=1}^{\infty}$ converges to x_k for each $k \in \mathbb{N}$. Defining the functionals ϕ_m as above, we see that $\phi_m(x^{(n)}) \to x_m$ for all $m \in \mathbb{N}$. Since $e^{(k)}$, $k \in \mathbb{N}$ is fundamental in ℓ_p^*, it follows by (1.9.7) that $\{x^{(n)}\}_{n=1}^{\infty}$ converges weakly to x. This completes the proof. □

Theorem 1.9.13. *Let S be a compact Hausdorff space, and let $x, \{x_n\}_{n=1}^{\infty} \in \mathcal{C}(S)$. Then $\{x_n\}_{n=1}^{\infty}$ converges weakly to x if and only if $\|x_n\|_\infty$ is bounded and $x_n(s) \to x(s)$ for each $s \in S$.*

Theorem 1.9.14 (Schur's Lemma)**.** *In the normed linear space ℓ_1, a sequence converges weakly if and only if it converges in norm.*

Definition 1.9.15 (Weakly Sequentially Closed)**.** *A subset F in a normed linear space X is said to be* **weakly sequentially closed** *if the weak limit of any weakly convergent sequence in F is also in F.*

Proposition 1.9.16. *Let X be a normed linear space. If a subset F of X is weakly sequentially closed, then F is closed.*

Proof. Let F be a weakly sequentially closed subset of a normed linear space X. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in F converging to a point $x \in X$. But since $\{x_n\}_{n=1}^{\infty}$ also converges weakly, we have $x \in F$ by the assumption. Hence, F is closed. □

Example 1.9.17. *The converse of* (1.9.16) *is not true in general, for consider $F := \{x \in c_0 : \|x\| = 1\}$. Then F is closed and we have shown that the standard basis vectors $\{e_n\}_{n=1}^{\infty}$ in F converge weakly to zero, but $0 \notin F$.*

Theorem 1.9.18. *A subspace of a normed linear space is closed if and only if it is weakly sequentially closed.*

Proof. We need only show that closure in the norm topology implies closure in the weak topology.

Let X be a normed linear space and let F be a closed subspace of X. Let $\{y_n\}_{n=1}^{\infty}$ converge weakly to a point $y \in X$. By contradiction, suppose that $y \notin F$. Then, since F is closed, we have $\operatorname{dist}(y, F) > 0$. Thus, by Corollary (1.6.18), there exists a continuous linear functional $\phi \in F^{\perp}$ such that $\phi(y) = 1$, but

$$\phi\left(\lim_{n\to\infty} y_n\right) = \lim_{n\to\infty} \phi(y_n) = 0,$$

a contradiction.

Hence, F is weakly sequentially closed. $\square$

More generally, a convex set is closed if and only if it is weakly sequentially closed.

Definition 1.9.19 (Weakly Sequentially Continuous)**.** *Let X and Y be normed linear spaces. A map $T : X \to Y$ is said to be* **weakly sequentially continuous** *at $x \in X$ if for all $\phi \in Y^*$ and every sequence $\{x_n\}_{n=1}^{\infty}$ converging to x in a weak sense, $\{(\phi \circ T)(x_n)\}_{n=1}^{\infty}$ converges to $(\phi \circ T)(x)$.*

Theorem 1.9.20. *Let X and Y be normed linear spaces. A linear mapping $T : X \to Y$ is continuous if and only if it is weakly sequentially continuous.*

Proof. First suppose that $T : X \to Y$ is continuous, and let $\{x_n\}_{n=1}^{\infty}$ be a sequence in X converging weakly to $x \in X$. Note that for all $\phi \in Y^*$, $\phi \circ T \in X^*$. Hence,

$$(\phi \circ T)\left(\lim_{n\to\infty} x_n\right) = \phi\left(\lim_{n\to\infty} (Tx_n)\right) = \phi(Tx),$$

which shows that T is weakly sequentially continuous.

For the converse, suppose by contradiction that T is weakly sequentially continuous but not continuous. Then T is unbounded, so that there exists a sequence $\{x_n\}_{n=1}^{\infty}$ such that $\|x_n\| \leq 1$ for all $n \in \mathbb{N}$, but $\|Tx_n\| \geq n^2$. Define a sequence $\{y_n\}_{n=1}^{\infty}$ by $y_n := x_n/n$ for all $n \in \mathbb{N}$. Then $\{y_n\}_{n=1}^{\infty}$ converges to zero, but $\|Ty_n\| > n$. Thus $\{Ty_n\}_{n=1}^{\infty}$ cannot converge weakly, for it is unbounded. $\square$

In the adjoint X^*, the concept of weak convergence is also available. That is, $\phi_n \rightharpoonup \phi$ if and only if $\Phi(\phi_n) \to \Phi(\phi)$ for each $\Phi \in X^{**}$.

Definition 1.9.21 (Weak* Convergence)**.** *Let X be a normed linear space. A sequence $\{\phi_n\}_{n=1}^{\infty}$ is said to* **converge in the weak* sense** *if $\{\phi_n(x)\}_{n=1}^{\infty}$ converges to $\phi(x)$ for all $x \in X$.*

Definition 1.9.22 (Separable)**.** *A normed linear space X is said to be* **separable** *if X contains a countable dense subset.*

Theorem 1.9.23. *Let X be a separable normed linear space and let $\{\phi_n\}_{n=1}^{\infty}$ be a bounded sequence in X^*. Then there is a subsequence $\{\phi_{n_k}\}_{k=1}^{\infty}$ that converges in the weak* sense to an element $\phi \in X^*$.*

Proof. Since X is separable, there exists a countable dense subset, say $\{x_1, x_2, \dots\}$. Since $\{\phi_n\}_{n=1}^{\infty}$ is bounded, it follows that the sequence $\{\phi_n(x_1)\}_{n=1}^{\infty}$ is also bounded. By the Bolzano–Weierstrass Theorem, there exists $I_1 \subseteq \mathbb{N}$ such that the subsequence $\{\phi_n(x_1)\}_{n\in I_1}$ converges. Similarly, the sequence $\{\phi_n(x_2)\}_{n\in\mathbb{N}}$ is bounded, so that there exists $I_2 \subseteq I_1$

such that the subsequence $\{\phi_n(x_2)\}_{n\in I_2}$ converges. Continuing in this fashion, we obtain sequences

$$\mathbb{N} \supseteq I_1 \supseteq I_2 \supseteq \ldots$$

and an array of functionals

$$\begin{matrix} \phi_{n_{11}}(x_1) & \phi_{n_{12}}(x_1) & \phi_{n_{13}}(x_1) & \ldots \\ \phi_{n_{21}}(x_2) & \phi_{n_{22}}(x_2) & \phi_{n_{23}}(x_2) & \ldots \\ \phi_{n_{31}}(x_3) & \phi_{n_{32}}(x_3) & \phi_{n_{33}}(x_3) & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{matrix}$$

such that $\lim_{k\to\infty} \phi_{n_{jk}}(x_j)$ exists for each $j \in \mathbb{N}$.

By Cantor's diagonalization process, define n_k to be the $k-$th element of I_k. We first note that $\lim_{i\to\infty} \phi_{n_i}(x_k)$ exists for each $k \in \mathbb{N}$, because $\lim_{n\in I_k} \phi_n(x_k)$ exists by the construction, and if $i \geq k$, then $i \in I_i \subseteq I_k$.

Fix $\epsilon > 0$ and $x \in X$. Since the sequence $\{\phi_n\}_{n=1}^{\infty}$ is bounded, there exists $M > 0$ such that $\|\phi_n\| \leq M$ for all $n \in \mathbb{N}$. Since $\{x_k\}_{k\in\mathbb{N}}$ is dense in X, there exists $k \in \mathbb{N}$ such that $\|x_k - x\| < \frac{\epsilon}{3M}$. Moreover, since $\lim_{i\to\infty} \phi_{n_i}(x_k)$ exists for each $k \in \mathbb{N}$, there exists a positive integer N such that for all $l, m \geq N$, we have $|\phi_{n_m}(x_k) - \phi_{n_l}(x_k)| < \epsilon/3$. Hence, for all $m, l \geq N$,

$$\begin{aligned} |\phi_{n_m}(x) - \phi_{n_l}(x)| &\leq |\phi_{n_m}(x) - \phi_{n_m}(x_k)| + |\phi_{n_m}(x_k) - \phi_{n_l}(x_k)| + |\phi_{n_l}(x_k) - \phi_{n_l}(x)| \\ &< \|\phi_{n_m}\|\|x - x_k\| + \epsilon/3 + \|\phi_{n_l}\|\|x_k - x\| \\ &< M\left(\frac{\epsilon}{3M}\right) + \epsilon/3 + M\left(\frac{\epsilon}{3M}\right) \\ &= \epsilon. \end{aligned}$$

This shows $\{\phi_{n_i}(x)\}_{i\in I}$ is Cauchy in $\mathbb{R}$ for all $x \in X$. Since $\mathbb{R}$ is complete, $\lim_{i\to\infty} \phi_{n_i}(x)$ exists for all $x \in X$, say $\lim_{i\to\infty} \phi_{n_i}(x) =: \phi(x)$.

We now show that $\phi \in X^*$. First, we see that ϕ is linear, for

$$\begin{aligned} \phi(\alpha x + \beta y) &= \lim_{i\to\infty} \phi_{n_i}(\alpha x + \beta y) = \alpha \lim_{i\to\infty} \phi_{n_i}(x) + \beta \lim_{i\to\infty} \phi_{n_i}(y) \\ &= \alpha\phi(x) + \beta\phi(y). \end{aligned}$$

Next, we show that ϕ is continuous. It suffices to show that ϕ is continuous at zero. Let $\{z_n\}_{n=1}^{\infty}$ be a sequence in X converging to zero. Observe

$$\left|\phi\left(\lim_{n\to\infty} z_n\right)\right| \leq \left\|\left(\lim_{i\to\infty} \phi_{n_i}\right)\right\| \left\|\left(\lim_{n\to\infty} z_n\right)\right\| \leq M \left\|\left(\lim_{n\to\infty} z_n\right)\right\| = 0.$$

This shows $\phi \in X^*$.

Lastly, to see that $\{\phi_{n_i}\}_{i\in I}$ converges to ϕ in the weak* sense, letting $l \to \infty$ in the above inequalities gives for all $m \geq N$

$$|\phi_{n_m}(x) - \phi_{n_l}(x)| < \epsilon.$$

Hence, $\{\phi_{n_i}\}_{i\in I}$ converges weak* to ϕ.

This completes the proof. $\square$

1.10. **Reflexive Spaces.** Let X be a Banach space. We may embed X isomorphically and isometrically as a subspace of X^{**}.

Definition 1.10.1 (Natural Embedding)**.** *Let X be a Banach sapce. The mapping $J : X \to X^{**}$ defined by*

$$(Jx)(\phi) = \phi(x)$$

for all $\phi \in X^*$ and $x \in X$ is called the **natural embedding** *of X into X^{**}.*

Proposition 1.10.2. *Let X be a Banach space. The natural mapping $J : X \to X^{**}$ is a linear isometry.*

Proof. Let $x, y \in X$, $\alpha, \beta \in \mathbb{R}$, and $\phi \in X^*$ be arbitrary. Then

$$(J(\alpha x + \beta y))(\phi) = \phi(\alpha x + \beta y) = \alpha\phi(x) + \beta\phi(y),$$

so that J is linear.

To see that J is isometric, we first note that

$$\|Jx\| = \sup_{\|\phi\|=1} \|(Jx)(\phi)\| = \sup_{\|\phi\|=1} |\phi(x)| \leq \sup_{\|\phi\|=1} \|\phi\|\|x\| = \|x\|.$$

For the reverse inequality, we have by (1.6.19) that there exists a continuous linear functional $\psi \in X^*$ such that $\|\psi\| = 1$ and $\psi(x) = \|x\|$. Hence,

$$\|Jx\| = \sup_{\|\phi\|=1} \|(Jx)(\phi)\| = \sup_{\|\phi\|=1} |\phi(x)| \geq |\psi(x)| = \|x\|,$$

which completes the proof. □

Definition 1.10.3 (Reflexive Space)**.** *Let X be a Banach space. If the natural map $J : X \to X^{**}$ is surjective, then X is called a* **reflexive space***.*

Note that if X is reflexive, then it is isometrically isomorphic to X^{**}. the converse is false, that is, there exist X and X^{**} isometrically isomorphic but X is not reflexive.

Theorem 1.10.4. *Each space ℓ_p, $1 < p < \infty$, is reflexive.*

Proof. If $1/p + 1/q = 1$, then $\ell_p = \ell_q^*$ and $\ell_q^* = \ell_p$ by (1.9.12). Evidently $\ell_p = \ell_p^{**}$. It suffices to check that J is the isometry. Let $A : \ell_p \to \ell_q^*$ and $B : \ell_q \to \ell_p^*$ be the isometries defined by

$$(Ax)(y) = \sum_{n=1}^{\infty} x_n y_n$$

for all $x \in \ell_p$, $y \in \ell_q$, and

$$(By)(x) = \sum_{n=1}^{\infty} x_n y_n$$

for all $x \in \ell_q$, $y \in \ell_p$.

Define $B^* : \ell_p^{**} \to \ell_q^*$ by

$$B^*(\phi) := \phi \circ B$$

for all $\phi \in \ell_p^{**}$. Then B^* is an isometric isomorphism of ℓ_p^{**} onto ℓ_q^*. Consequently, B^{*-1} is an isometric isomorphism of ℓ_p onto ℓ_p^{**}. We show that $B^{*-1} = J$. Observe that for all $x \in \ell_p$, $y \in \ell_q$,

$$\begin{aligned} B^{*-1}Ax = Jx &\iff Ax = B^*Jx \\ &\iff (Ax)(y) = (B^*Jx)(y) \end{aligned}$$

$$\iff (Ax)(y) = (Jx)(By)$$
$$\iff (Ax)(y) = (By)(x),$$

where

$$(Ax)(y) = \sum_{n=1}^{\infty} x_n y_n = (By)(x).$$

This completes the proof. □

Theorem 1.10.5. *A closed linear subspace in a reflexive Banach space is reflexive.*

Proof. Let Y be a closed subspace of a reflexive Banach space X. Let $J : X \to X^{**}$ be the natural map, and define $R : X^* \to Y^*$ by $R\phi := \phi|_Y$. Let $f \in Y^{**}$. Put $y = J^{-1}(f \circ R)$. We show that $y \in Y$.

By contradiction, suppose that $y \notin Y$. By (1.6.18), there exists a continuous linear functional $\phi \in X^*$ such that $\phi(y) = 1$ and $\phi(Y) = 0$. Then $R\phi = 0$ and thus

$$\phi(y) = \phi(J^{-1}(f \circ R)) = (f \circ R)(\phi) = 0,$$

a contradiction.

Lastly, we show that for all $\psi \in Y^*$, $f(\psi) = \psi(y)$. Let $\widetilde{\psi}$ be a Hahn–Banach extension of ψ in X^*. Then $\psi = R\widetilde{\psi}$ and

$$f(\psi) = f(R\widetilde{\psi}) = (f \circ R)(\widetilde{\psi}) = (Jy)(\widetilde{\psi}) = \widetilde{\psi}(y) = \psi(y),$$

which completes the proof. □

Theorem 1.10.6. *A Banach space is reflexive if and only if its conjugate space is reflexive.*

Proof. First suppose that X is a reflexive Banach space. Then the natural embedding $J : X \to X^{**}$ is surjective. Let $\Phi \in X^{***}$ and define $\phi \in X^*$ by $\phi := \Phi \circ J$. Then for any $f \in X^{**}$ we have $f = Jx$ for some $x \in X$, and consequently

$$f(\phi) = (Jx)(\phi) = \phi(x) = (\Phi \circ J)(x) = \Phi(Jx) = \Phi(f).$$

Thus Φ is the image of ϕ under the natural map of X^* into X^{***}, which shows that X^* is reflexive.

For the converse, assume that X^* is reflexive. By the above argument, X^{**} is reflexive. But $J(X)$ is a closed subspace in X^{**}, and by (1.10.5), $J(X)$ is reflexive. Since X is isometrically isomorphic to $J(X)$, X is reflexive.

This completes the proof. □

The following two theorems are presented without proof.

Theorem 1.10.7 (Eberlein–Smulyan Theorem)**.** *A Banach space is reflexive if and only if its unit ball is weakly sequentially compact.*

Theorem 1.10.8. *A Banach space X is reflexive if and only if each continuous linear functional on X attains its supremum on the unit ball of X.*

One application of the second adjoint occurs in the process of *completion.* Given a normed linear space X that is not complete, we want to embed it linearly and isometrically as a dense set in a Banach space. We call such a Banach space a *completion* of X.

Definition 1.10.9 (Completion)**.** *Let X be a normed linear space that is not complete. If X can be embedded as a dense set in a Banach space $\mathcal{X}$ under a linear isometry, we call $\mathcal{X}$ the completion of X.*

Note that if X is a normed linear space, we can embed it, using the natural map J, into its double adjoint X^{**}. Recall from (1.6.30) that X^{**} is automatically complete. Hence, $\overline{J(X)}$ can be regarded as a completion of X. It may be proved that all completions of X are isometrically isomorphic to each other.

Consider the Lebesgue spaces $L^p[a,b]$. Denote by $\mathcal{C}[a,b]$ the space of all continuous real–valued functions on the interval $[a,b]$. For $1 \leq p < \infty$, we introduce the norm

$$\|x\|_p := \left\{ \int_a^b |x(t)|^p \, dt \right\}^{1/p}.$$

In this equation, the integration is with respect to the Riemann integral. Then the space $\mathcal{C}[a,b]$, endowed with this norm, is denoted by $\mathcal{C}_p[a,b]$, and is not complete. Its completion is $L^p[a,b]$. Thus, if J is the natural map of $\mathcal{C}_p[a,b]$ into its double adjoint, then

$$L^p[a,b] = \overline{J(\mathcal{C}_p[a,b])}.$$

2. Hilbert Spaces

2.1. **Geometry.** Hilbert spaces are a special type of Banach space. The distinguishing characteristic is that the parallelogram law is assumed to hold:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Definition 2.1.1 (Inner Product)**.** *Let X be a complex linear space. An* **inner product** *is a function $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$ such that for all $x, y, z \in \mathbb{C}$ and $\alpha \in \mathbb{C}$, the following properties hold:*

(1) $\langle x, y \rangle \in \mathbb{C}$;
(2) $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
(3) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$;
(4) $\langle x, x \rangle > 0$, $x \neq 0$;
(5) $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$.

Definition 2.1.2 (Inner Product Space)**.** *An* **inner product space** *is a pair $(X, \langle \cdot, \cdot \rangle)$ where X is a linear space and $\langle \cdot, \cdot \rangle$ is an inner product on X.*

Inner product spaces may also be called "pre–Hilbert spaces." Occasionally we also consider real inner product spaces and real Hilbert spaces.

Example 2.1.3. *Let $X = \mathbb{C}^n$. Let $x = (x_1, \ldots, x_n)^\top, y = (y_1, \ldots, y_n)^\top \in \mathbb{C}^n$. We define*

$$\langle x, y \rangle := \sum_{j=1}^{n} x_j \overline{y}_j.$$

Example 2.1.4. *Let $X = \mathcal{C}[[0,1], \mathbb{C}]$. Let $x, y \in \mathcal{C}[[0,1], \mathbb{C}]$. We define*

$$\langle x, y \rangle := \int_0^1 x(t)\overline{y(t)}\, dt.$$

Proposition 2.1.5. *Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. Then for all $x, y, z \in X$ and $\alpha \in \mathbb{C}$,*

(1) $\langle x+y, x+y \rangle = \langle x, x \rangle + 2\mathrm{Re}\, \langle x, y \rangle + \langle y, y \rangle$;
(2) $\langle x, \alpha y \rangle = \overline{\alpha} \langle x, y \rangle$;
(3) $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$;
(4) $\left\langle \sum_{j=1}^{n} x_j, y \right\rangle = \sum_{j=1}^{n} \langle x_j, y \rangle$.

Proof. We show (1), (2), and (3).

(1).

$$\begin{aligned}
\langle x+y, x+y \rangle &= \langle x, x+y \rangle + \langle y, x+y \rangle = \overline{\langle x+y, x \rangle} + \overline{\langle x+y, y \rangle} \\
&= \overline{\langle x, x \rangle} + \overline{\langle y, x \rangle} + \overline{\langle x, y \rangle} + \overline{\langle y, y \rangle} \\
&= \langle x, x \rangle + \langle y, y \rangle + \langle x, y \rangle + \overline{\langle y, x \rangle} \\
&= \langle x, x \rangle + \langle y, y \rangle + 2\mathrm{Re}\, \langle x, y \rangle .
\end{aligned}$$

(2).

$$\langle x, \alpha y \rangle = \overline{\langle \alpha y, x \rangle} = \overline{\alpha}\overline{\langle y, x \rangle} = \overline{\alpha} \langle x, y \rangle .$$

(3).

$$\langle x, y+z \rangle = \overline{\langle y+z, x \rangle} = \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle .$$

∎

Definition 2.1.6 (Norm)**.** *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space. We define the* **induced norm** $\|\cdot\|$ *by*

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

Theorem 2.1.7. *Let $(X, \langle\cdot,\cdot\rangle)$ be an inner product space, let $x, y \in X$, and let $\alpha \in \mathbb{C}$. The induced norm $\|\cdot\|$ has the following properties:*

(1) $\|x\| > 0$, $x \neq 0$;
(2) $\|\alpha x\| = |\alpha|\|x\|$;
(3) $|\langle x, y\rangle| \leq \|x\|\|y\|$ *(Cauchy–Schwarz inequality);*
(4) $\|x+y\| \leq \|x\| + \|y\|$ *(triangle inequality);*
(5) $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$ *(parallelogram equality);*
(6) If $\langle x, y\rangle = 0$*, then* $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ *(Pythagorean law).*

Before giving the proof, note that (1), (2), and (4) show that the induced norm indeed defines a norm over X.

Proof. We show (3), (4), and (5). Note that (6) follows from Proposition ((2.1.5), 1).

For (3), we see that for any $\lambda \in \mathbb{C}$,

$$0 \leq \langle x - \lambda y, x - \lambda y\rangle = \langle x, x\rangle - \overline{\lambda}\langle x, y\rangle - \lambda\langle y, x\rangle + |\lambda|^2\langle y, y\rangle .$$

If $y = 0$, the result follows trivially, so assume that $y \neq 0$. Define $\lambda := \frac{\langle y,x\rangle}{\langle y,y\rangle}$. Then

$$\begin{aligned} 0 &\leq \|x\|^2 - \frac{\overline{\langle y, x\rangle}}{\langle y, y\rangle}\langle x, y\rangle - \frac{\langle y, x\rangle}{\langle y, y\rangle}\langle y, x\rangle + \left|\frac{\langle y, x\rangle}{\langle y, y\rangle}\right|^2 \|y\| \\ &= \|x\|^2 - 2\frac{|\langle y, x\rangle|^2}{\|y\|^2} + \frac{|\langle y, x\rangle|^2}{\|y\|^2} \\ &= \|x\|^2 - \frac{|\langle y, x\rangle|^2}{\|y\|^2}. \end{aligned}$$

Since $|\langle y, x\rangle|^2 = |\langle x, y\rangle|^2$, this shows

$$|\langle x, y\rangle| \leq \|x\|\|y\|.$$

For (4), we apply (3) as follows:

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

Lastly, for (5) we have

$$\begin{aligned} \|x+y\|^2 + \|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle + \langle x, x\rangle + \langle x, -y\rangle + \langle -y, x\rangle + \langle y, y\rangle \\ &= 2\|x\|^2 + 2\|y\|^2 + \langle x, y\rangle + \langle y, x\rangle - \langle x, y\rangle - \langle y, x\rangle \\ &= 2\|x\|^2 + 2\|y\|^2. \end{aligned}$$

□

The parallelogram law is so named because it states that the sum of the squares of the four sides of a parallelogram is equal to the sum of the squares of the two diagonals.

Lemma 2.1.8. *Let* $(X, \langle \cdot, \cdot \rangle)$ *be an inner product space and let* $x, y \in X$. *Then*

(1) $x = 0$ *if and only if* $\langle x, v \rangle = 0$ *for all* $v \in X$;
(2) $x = y$ *if and only if* $\langle x, v \rangle = \langle y, v \rangle$ *for all* $v \in X$;
(3) $\|x\| = \sup_{\|v\|=1} |\langle x, v \rangle|$.

Proof. For (1), first note that if $\langle x, v \rangle = 0$ for all $v \in X$, then in particular $\langle x, x \rangle = 0$, from which it follows $x = 0$. Conversely, if $x = 0$, then it follows by Axiom 3 that $\langle x, v \rangle = \langle 0, v \rangle = 0$ for all $v \in X$.

The condition $x = y$ is equivalent to $x - y = 0$, to $\langle x - y, v \rangle = 0$ for all $v \in X$, and thus to $\langle x, v \rangle = \langle y, v \rangle$ for all $v \in X$.

Lastly, if $\|v\| = 1$, then $|\langle x, v \rangle| \leq \|x\| \|v\| = \|x\|$ by the Cauchy–Schwarz inequality. For the reverse inequality, first note that if $x = 0$, then $\|x\| \leq |\langle x, v \rangle|$ for all $v \in X$. Otherwise, let $v := x/\|x\|$. Then $\|v\| = 1$ and

$$|\langle x, v \rangle| = |\langle x, x/\|x\| \rangle| = \frac{1}{\|x\|} \langle x, x \rangle = \|x\|.$$

□

Definition 2.1.9 (Hilbert Space)**.** *A* **Hilbert space** *is an inner product space that is complete.*

Example 2.1.10. *The space* $\mathcal{C}[[0,1], \mathbb{C}]$ *with inner product*

$$\langle x, y \rangle := \int_0^1 x(t)\overline{y(t)}\, dt$$

is not complete. Consider the functions $\{x_n\}_{n=1}^{\infty}$ *defined by*

$$x_n(t) := \begin{cases} 0, & 0 \leq t \leq \frac{1}{2} - \frac{1}{n}, \\ nt + (1 - n/2), & \frac{1}{2} - \frac{1}{n} < t < \frac{1}{2}, \\ 1, & t \geq \frac{1}{2}. \end{cases}$$

Then, for all $n, m \in \mathbb{N}$ such that $m > n$,

$$\|x_n - x_m\|^2 = \int_0^1 |x_n(t) - x_m(t)| \leq \int_{1/2-1/n}^{1/2} dt = t|_{1/2-1/n}^{1/2} = \frac{1}{n} \to 0 \quad \text{as } n \to \infty.$$

On the other hand, this sequence converges pointwise to

$$x(t) := \begin{cases} 0, & t < 1/2, \\ 1, & t \geq 1/2, \end{cases}$$

which is not in $\mathcal{C}[[0,1], \mathbb{C}]$.

Example 2.1.11. *We write* $L^2[a,b]$ *for the set of all complex–valued Lebesgue square integrable functions on* $[a,b]$, *that is,*

$$\int_a^b |x|^2 \, d\mu < \infty$$

for all $x \in L^2[a,b]$, with μ the Lebesgue measure. Put

$$\langle x, y\rangle := \int_a^b x(t)\overline{y(t)}\ d\mu.$$

Then $L^2[a,b]$ is complete and thus is a Hilbert space.

Recall that in $L^2[a,b]$ two functions f and g are equivalent if and only if they differ only on a set of measure zero. The elements of $L^2[a,b]$ are thus not functions, but equivalence classes of functions.

Example 2.1.12. *Let $(X, \mathcal{S}, \mu)$ be any measure space. The notation $L^2(X)$ then denotes the space of all complex–valued $\mathcal{S}-$measurable functions f on X such that $\int |f|^2\ d\mu < \infty$. In $L^2(X)$, we define*

$$\langle f, g\rangle := \int f(t)\overline{g(t)}\ d\mu(t).$$

Then $L^2(S)$ is a Hilbert space.

Example 2.1.13. *The space ℓ_2 of all complex sequences $x = \{x_n\}_{n=1}^{\infty}$ such that $\sum_{n=1}^{\infty} |x_n|^2 < \infty$ with inner product*

$$\langle x, y\rangle := \sum_{n=1}^{\infty} x_n\overline{y_n}$$

is a Hilbert space. To see this, take $X = \mathbb{N}$ in the previous example and take μ to be the counting measure.

Theorem 2.1.14. *Let K be a nonempty, closed, convex set in a Hilbert space X. Then to each $x \in X$ there corresponds a unique point $y \in K$ closest to x, that is,*

$$\|x - y\| = \operatorname{dist}(x, K) := \inf_{v\in K} \|x - v\|.$$

Proof. Fix $x \in X$, and put $\alpha := \operatorname{dist}(x, K)$. By definition of an infimum, there exists a sequence $\{y_n\}_{n=1}^{\infty}$ in K such that $\|x - y_n\| \to \alpha$ as $n \to \infty$. By the convexity of K, note that $\frac{1}{2}(y_n + y_m) \in K$ for each $n, m \in \mathbb{N}$. Thus, by the assumption, $\|\frac{1}{2}(y_n + y_m) - x\| \geq \alpha$. By the parallelogram law, observe

$$\begin{aligned}
\|y_n - y_m\|^2 &= \|(y_n - x) - (y_m - x)\|^2 \\
&= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - \|y_n + y_m - 2x\|^2 \\
&= 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\|\frac{1}{2}(y_n + y_m) - x\|^2 \\
&\leq 2\|y_n - x\|^2 + 2\|y_m - x\|^2 - 4\alpha^2,
\end{aligned}$$

which tends to zero as $n \to \infty$. Thus $\{y_n\}_{n=1}^{\infty}$ is a Cauchy sequence. Since X is complete, $\{y_n\}_{n=1}^{\infty}$ converges to y for some $y \in X$. Moreover, since K is closed, $y \in K$. By continuity of the induced norm,

$$\|x - y\| = \|x - \lim_{n\to\infty} y_n\| = \lim_{n\to\infty} \|x - y_n\| = \alpha = \operatorname{dist}(x, K).$$

To show uniqueness, suppose that $y_1, y_2 \in K$ are such that

$$\|x - y_1\| = \operatorname{dist}(x, K) = \|x - y_2\|.$$

By the above calculation and the parallelogram law we have

$$\|y_1 - y_2\| \leq 2\|y_1 - x\|^2 + 2\|y_2 - x\|^2 - 4\alpha^2 = 0,$$

which shows $y_1 = y_2$.

This completes the proof. $\square$

Definition 2.1.15 (Orthogonal)**.** *Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space. We say that $x, y \in X$ are* **orthogonal** *if $\langle x, y \rangle = 0$, and we write $x \perp y$.*

Remark. *Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space.*

(1) If $Y \subseteq X$, then $x \perp Y$ means that $\langle x, y \rangle = 0$ for all $y \in Y$.
(2) If $U, V \subseteq X$, then $U \perp V$ means that $\langle u, v \rangle = 0$ for all $u \in U$ and $v \in V$.

Theorem 2.1.16. *Let Y be a subspace in an inner product space X. Let $x \in X$ and $y \in Y$. Then $x - y \perp Y$ if and only if y is the unique point in Y closest to x.*

Proof. First suppose that $x - y \perp Y$. Let $v \in Y$, and note that $y - v \in Y$. Then by the Pythagorean law,

$$\|x - v\|^2 = \|(x - y) + (y - v)\|^2 = \|x - y\|^2 + \|y - v\|^2 \geq \|x - y\|^2.$$

Assume now that y is the unique point in Y closest to x. Then for any $v \in Y$ and $\lambda \in \mathbb{C}$,

$$\begin{aligned}
0 &\leq \|x - (y + \lambda v)\|^2 - \|x - y\|^2 = \|(x - y) - \lambda v\|^2 - \|x - y\|^2 \\
&= \langle x - y - \lambda v, x - y - \lambda v \rangle - \langle x - y, x - y \rangle \\
&= \langle x - y, x - y - \lambda v \rangle + \langle -\lambda v, x - y - \lambda v \rangle - \langle x - y, x - y \rangle \\
&= \langle x - y, x - y \rangle + \langle x - y, -\lambda v \rangle + \langle -\lambda v, x - y \rangle + \langle -\lambda v, -\lambda v \rangle - \langle x - y, x - y \rangle \\
&= -2\mathrm{Re}\, \langle x - y, \lambda v \rangle + |\lambda|^2 \|v\|^2,
\end{aligned}$$

that is,

$$2\mathrm{Re}(\overline{\lambda}\, \langle x - y, v \rangle) \leq |\lambda|^2 \|v\|^2.$$

Suppose by contradiction that $\langle x - y, v \rangle \neq 0$. Then $v \neq 0$, so we may define

$$\lambda := \frac{\langle x - y, v \rangle}{\|v\|^2}.$$

But then

$$2\mathrm{Re}\left(\frac{|\langle x - y, v \rangle|^2}{\|v\|^2} \right) \leq \frac{|\langle x - y, v \rangle|^2}{\|v\|^2},$$

a contradiction. $\square$

Definition 2.1.17 (Orthogonal Complement)**.** *Let Y be a subset in an inner product space X. Then the* **orthogonal complement** *of Y is given by*

$$Y^{\perp} := \{x \in X : \langle x, y \rangle = 0 \textit{ for all } y \in Y\}.$$

Theorem 2.1.18. *If Y is a closed subspace of a Hilbert space X, then $X = Y \oplus Y^{\perp}$.*

Proof. We first show that $Y^{\perp}$ is a subspace. Let $v_1, v_2 \in Y^{\perp}$, and let $\alpha, \beta \in \mathbb{C}$. Then for all $y \in Y$,

$$\begin{aligned}\langle y, \alpha v_1 + \beta v_2 \rangle &= \langle y, \alpha v_1 \rangle + \langle y, \beta v_2 \rangle \\ &= \overline{\alpha} \langle y, v_1 \rangle + \overline{\beta} \langle y, v_2 \rangle \\ &= 0.\end{aligned}$$

Thus $\alpha v_1 + \beta v_2 \in Y^{\perp}$.

Next, we show that $Y \cap Y^{\perp} = \{0\}$. Suppose that $x \in Y \cap Y^{\perp}$. Then $\langle x, x \rangle = 0$, so that $x = 0$.

Finally, we show that $X \subseteq Y \oplus Y^{\perp}$. Let $x \in X$ be arbitrary. By Theorem (2.1.14), there exists a unique point $y \in Y$ closest to x. Moreover, by Theorem (2.1.16), $x - y \in Y^{\perp}$. Noting that $x = y + (x - y) \in Y \oplus Y^{\perp}$, this completes the proof. □

Theorem 2.1.19. *If the parallelogram law holds in a normed linear space X, then X is an inner product space. That is, an inner product may be defined on X such that $\langle x, x \rangle = \|x\|^2$ for all $x \in X$.*

Proof. We define the inner product $\langle \cdot, \cdot \rangle$ as follows: for all $x, y \in X$,

$$4 \langle x, y \rangle := \|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2.$$

Clearly $\langle \cdot, \cdot \rangle \in \mathbb{C}$.

By the construction,

$$4\mathrm{Re} \langle x, y \rangle = \|x + y\|^2 - \|x - y\|^2.$$

By the parallelogram equality, we obtain

$$\begin{aligned}4\mathrm{Re} \langle u + v, w \rangle &= \|u + v + w\|^2 - \|u + v - w\|^2 \\ &= \Big(2\|u + w\|^2 + 2\|v\|^2 - \|u + w - v\|^2\Big) - \\ &\qquad \Big(2\|u\|^2 + 2\|v - w\|^2 - \|u - v + w\|^2\Big) \\ &= 2\|u + w\|^2 + 2\|v\|^2 - 2\|u\|^2 - 2\|v - w\|^2 \\ &= 2\|u + w\|^2 + \{\|v + w\|^2 + \|v - w\|^2 - 2\|w\|^2\} - \\ &\qquad \{\|u + w\|^2 + \|u - w\|^2 - 2\|w\|^2\} - 2\|v - w\|^2 \\ &= \|u + w\|^2 - \|u - w\|^2 + \|v + w\|^2 - \|v - w\|^2 \\ &= 4\mathrm{Re} \langle u, w \rangle + 4\mathrm{Re} \langle v, w \rangle .\end{aligned}$$

Putting iy in place of y in the definition of $\langle x, y \rangle$ gives

$$4 \langle x, iy \rangle = \|x + iy\|^2 - \|x - iy\|^2 + i\|x - y\|^2 - i\|x + y\|^2 = -4i \langle x, y \rangle .$$

Thus we have

$$\begin{aligned}\mathrm{Im} \langle u + v, w \rangle &= -\mathrm{Re} i \langle u + v, w \rangle = \mathrm{Re} \langle u + v, iw \rangle \\ &= \mathrm{Re} \langle u, iw \rangle + \mathrm{Re} \langle v, iw \rangle = -\mathrm{Re} i \langle u, w \rangle - \mathrm{Re} i \langle v, w \rangle \\ &= \mathrm{Im} \langle u, w \rangle + \mathrm{Im} \langle v, w \rangle .\end{aligned}$$

This shows

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$$

for all $u, v, w \in X$.

By induction, we have $\langle nx, y\rangle = n\langle x, y\rangle$ for all $n \in \mathbb{N}$. From this it follows, for any $n, m \in \mathbb{N}$, that

$$\left\langle \frac{n}{m}x, y\right\rangle = \frac{n}{m}m\left\langle \frac{x}{m}, y\right\rangle = \frac{n}{m}\langle x, y\rangle .$$

By continuity, we obtain $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$ for any $\lambda \geq 0$. By the construction, we verify that

$$\langle -x, y\rangle = -\langle x, y\rangle$$

and

$$\langle ix, y\rangle = i\langle x, y\rangle .$$

Hence,

$$\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$$

for all $\lambda \in \mathbb{C}$.

From the definition, we have

$$\begin{aligned} 4\langle x, x\rangle &= \|2x\|^2 + i\|x + ix\|^2 - i\|x - ix\|^2 \\ &= 4\|x\|^2 + i|1 + i|^2\|x\|^2 - i|1 - i|^2\|x\|^2 = 4\|x\|^2 > 0. \end{aligned}$$

Finally, we observe that

$$\begin{aligned} 4\langle y, x\rangle &= \|y + x\|^2 - \|y - x\|^2 + i\|y + ix\|^2 - i\|y - ix\|^2 \\ &= \|x + y\|^2 - \|x - y\|^2 + i\| - i(y + ix)\|^2 - i\|i(y - ix)\|^2 \\ &= \|x + y\|^2 - \|x - y\|^2 + i\|x - iy\|^2 - i\|x + iy\|^2 \\ &= 4\overline{\langle x, y\rangle}. \end{aligned}$$

This completes the proof. □

In any inner product space X, the angle between two nonzero vectors can be defined. Recall the Law of Cosines: In a triangle having sides a, b, c and angle θ opposite side c, we have

$$c^2 = a^2 + b^2 - 2ab\cos\theta.$$

Notice that when $\theta = \pi/2$, this equation gives the Pythagorean Theorem. In an inner product space, recall that

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\mathrm{Re}\langle x, y\rangle .$$

On the other hand, we want the Law of Cosines to hold:

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos\theta.$$

Thus we define $\cos\theta$ so that $\|x\|\|y\|\cos\theta = \mathrm{Re}\langle x, y\rangle$. Therefore,

$$\theta = \arccos\left(\frac{\mathrm{Re}\langle x, y\rangle}{\|x\|\|y\|}\right).$$

Note that this angle is well-defined, since

$$\left|\frac{\mathrm{Re}\langle x, y\rangle}{\|x\|\|y\|}\right| \leq \frac{\|x\|\|y\|}{\|x\|\|y\|} = 1$$

by the Cauchy-Schwarz inequality.

2.2. **Orthogonality and Bases.**

Definition 2.2.1 (Orthogonal Set). *A set V of vectors in an inner product space X is said to be* **orthogonal** *if $\langle x, y\rangle = 0$ for all $x, y \in V$, $x \neq y$.*

Recall that we write $x \perp y$ if $\langle x, y\rangle = 0$, $x \perp S$ if $\langle x, y\rangle = 0$ for all $y \in S$, And $U \perp V$ if $\langle x, y\rangle = 0$ for all $x \in U$ and $y \in V$.

Theorem 2.2.2 (Pythagorean Law). *If $\{x_1, x_2, \dots, x_n\}$ is a finite orthogonal set of n distinct elements in an inner product space, then*

$$\left\|\sum_{j=1}^{n} x_n\right\|^2 = \sum_{j=1}^{n} \|x_j\|^2.$$

Proof. By the assumption, $x_i \neq x_j$ if $i \neq j$, and, consequently,

$$\begin{aligned}\left\|\sum_{j=1}^{n} x_n\right\|^2 &= \left\langle \sum_{j=1}^{n} x_j, \sum_{i=1}^{n} x_i \right\rangle = \sum_{j=1}^{n} \left\langle x_j, \sum_{i=1}^{n} x_i \right\rangle = \sum_{j=1}^{n}\sum_{i=1}^{n} \langle x_j, x_i\rangle \\ &= \sum_{j=1}^{n} \langle x_j, x_j\rangle = \sum_{j=1}^{n} \|x_j\|^2,\end{aligned}$$

completing the proof. □

The Pythagorean Law has an important counterpart for orthogonal sets which are not finite. First, we need to consider what it means to sum the elements in an arbitrary subset V of X. We take the following definition.

Definition 2.2.3. *Let V be an infinite subset of an inner product space X. We say that the sum of the elements of V is s if and only if the following condition is true: For any $\epsilon > 0$, there exists a finite subset V_0 of V such that for every larger finite subset F of V we have*

$$\left|\sum_{x\in F} x - s\right| < \epsilon.$$

In other words, we partially order the finite subsets of V by inclusion. With each finite subset F of V we associate the sum $S(F)$ of all the elements in F. Then S is a net, that is, a function on a directed set. The limit of this net, if it exists, is the sum s of all the elements of V. More precisely, it is often called the *unordered* sum over V.

Theorem 2.2.4 (General Pythagorean Law). *Let $\{x_j\}_{j\in I}$ be an orthogonal sequence in a Hilbert space. The series $\sum_{j\in I} x_j$ converges if and only if $\sum_{j\in I} \|x_j\|^2 < \infty$. If $\sum_{j\in I} \|x_j\|^2 = \lambda < \infty$, then $\|\sum_{j\in I} x_j\|^2 = \lambda$, and the sum $\sum_{j\in I} x_j$ is independent of the ordering of the terms.*

Proof. Put $S_n := \sum_{j=1}^{n} x_j$ and $s_n := \sum_{j=1}^{n} \|x_j\|^2$.

By the finite version of the Pythagorean Law (2.2.2), we have for all $m > n$ that

$$\|S_m - S_n\|^2 = \left\|\sum_{j=n+1}^{m} x_j\right\|^2 = \sum_{j=n+1}^{m} \|x_j\|^2 = |s_m - s_n|.$$

Thus $\{S_n\}_{n\in I}$ is a Cauchy sequence in X if and only if $\{s_n\}_{n\in I}$ is a Cauchy sequence in $\mathbb{R}$. This establishes the first assertion of the theorem.

Next, assume that $\lambda < \infty$. By the Pythagorean Law, $\|S_n\|^2 = s_n$, and thus in the limit we have

$$\left\|\sum_{j\in I} x_j\right\|^2 = \left\|\lim_{j\in I}\sum_j x_j\right\|^2 = \lim_{j\in I}\left\|\sum_j x_j\right\|^2 = \lim_{j\in I}\sum_j \|x_j\|^2 = \lambda.$$

Let u be any rearrangement of the original series, say that $u = \sum_{j\in I} x_{k_j}$. Let $U_n = \sum_{i=1}^n x_{k_j}$. Since any rearrangement of any absolutely convergent series in $\mathbb{R}$ converges to the same limit, we have $\sum_{j\in I}\|x_{k_j}\|^2 = \lambda$. Thus, by the previous analysis, we have $U_n \to u$ and $\|u\|^2 = \lambda$. Moreover,

$$\langle U_n, S_m\rangle = \left\langle \sum_{j=1}^n x_{k_j}, \sum_{i=1}^m x_i\right\rangle = \sum_{j=1}^n\sum_{i=1}^m \|x_i\|^2\delta_{ik_j}.$$

Letting $n\to\infty$, we obtain

$$\langle u, S_m\rangle = \sum_{i=1}^m \|x_i\|^2.$$

Now, letting $m\to\infty$, we get $\langle u, x\rangle = \lambda$, where $x = \lim_{m\to\infty} S_m$. It follows that $x = u$, because

$$\|x-u\|^2 = \|x\|^2 - 2\mathrm{Re}\,\langle x, u\rangle + \|u\|^2 = \lambda - 2\lambda + \lambda = 0.$$

This completes the proof. □

Definition 2.2.5 (Orthonormal Set). *A set U in an inner product space X is said to be* **orthonormal** *if U is an orthogonal set and each $v\in U$ has unit length, that is, $\|v\| = 1$ for all $v\in U$.*

Note that if $\{v_i\}_{i\in I}$ is an orthogonal set of nonzero vectors, then $\{v_i/\|v_i\|\}_{i\in I}$ is an orthonormal set.

Theorem 2.2.6. *If $\{y_j\}_{j=1}^n$ is an orthonormal set in an inner product space X, and if $Y = \operatorname{span}\{y_j\}_{j=1}^n$, then for any $x\in X$, the point in Y closest to x is $\sum_{j=1}^n \langle x, y_i\rangle y_i$.*

Proof. Put $y := \sum_{j=1}^n \langle x, y_i\rangle y_i$. By Theorem (2.1.16), it suffices to show that $x - y \perp Y$. We show that $x-y$ is orthogonal to each basis vector y_k, $k = 1, 2, \ldots, n$. We have

$$\begin{aligned}\langle x-y, y_k\rangle &= \langle x, y_k\rangle - \left\langle \sum_{j=1}^n \langle x, y_j\rangle y_j, y_k\right\rangle = \langle x, y_k\rangle - \sum_{j=1}^n \langle x, y_j\rangle\langle y_j, y_k\rangle \\ &= \langle x, y_k\rangle - \sum_{j=1}^n \langle x, y_j\rangle \delta_{jk} = \langle x, y_k\rangle - \langle x, y_k\rangle \\ &= 0,\end{aligned}$$

which completes the proof. □

Definition 2.2.7 (Orthogonal Projection). *Let $\{y_1, y_2, \ldots, y_n\}$ be an orthonormal set in an inner product space X and let $x\in X$. The* **orthogonal projection** *of x onto $\operatorname{span}\{y_1, y_2, \ldots, y_n\}$ is the vector $\sum_{k=1}^n \langle x, y_k\rangle y_k$.*

Definition 2.2.8 (Fourier Coefficients)**.** *Let* $\sum_{k=1}^{n} \langle x, y_k \rangle y_k$ *be the orthogonal projection of* x *onto* $\operatorname{span}\{y_1, y_2, \ldots, y_n\}$. *The coefficients* $\langle x, y_k \rangle$ *are called the* **Fourier coefficients** *of* x *with respect to* $\{y_1, y_2, \ldots, y_n\}$.

Definition 2.2.9 (Orthogonal Projection)**.** *Let* $\{y_1, y_2, \ldots, y_n\}$ *be an orthonormal set in an inner product space* X. *The operator that produces an orthogonal projection* y *from* $x \in X$ *is called an* **orthogonal projection**.

Corollary 2.2.10. *If* x *is a point in the linear span of an orthonormal set* $\{y_1, y_2, \ldots, y_n\}$, *then* $x = \sum_{i=1}^{n} \langle x, y_i \rangle y_i$.

Proof. The result follows immediately from Theorem (2.2.6). □

Theorem 2.2.11 (Bessel's Inequality)**.** *If* $\{u_i : i \in I\}$ *is an orthonormal system in an inner product space* X, *then for every* $x \in X$,

$$\sum_{i \in I} |\langle x, u_i \rangle|^2 \leq \|x\|^2.$$

Proof. Let $J \subseteq I$ be a finite subset of I. Let $y = \sum_{j \in J} \langle x, u_j \rangle u_j$. Note that y is the orthogonal projection of x onto the subspace $Y := \operatorname{span}\{u_j : j \in J\}$. By Theorem (2.1.16), $x - y \perp Y$. Thus by the Pythagorean Law, we have

$$\|x\|^2 = \|(x-y)+y\|^2 = \|x-y\|^2 + \|y\|^2 \geq \|y\|^2 = \sum_{j \in J} \|\langle x, u_j \rangle u_j\|^2 = \sum_{j \in J} |\langle x, u_j \rangle|^2.$$

This proves the result for any finite set J of indices.

We use Zorn's Lemma to show that

$$\sum_{i \in I} |\langle x, u_i \rangle|^2 \leq \|x\|^2.$$

Consider the collection J of subsets of I such that

$$\sum_{j \in J} |\langle x, u_j \rangle|^2 \leq \|x\|^2.$$

Partially order the collection J by inclusion. Let C be a totally ordered subset in J. Put

$$J^* := \cup_{J \in C} J.$$

We show that J^* is an upper bound for C.

Let $S := \sum_{j \in J^*} |\langle x, u_j \rangle|^2$. If J^* is finite, then $S \leq \|x\|^2$ by the first part of the proof. Otherwise, suppose by contradiction that $S > \|x\|^2$. Choose $\epsilon < \frac{S - \|x\|^2}{2}$. By definition of an infinite sum, there exists a finite subset J_0 of J^* such that for any larger finite subset $J_0 \subset F \subset J^*$, we have

$$\left| \sum_{j \in F} |\langle x, u_j \rangle|^2 - S \right| < \epsilon.$$

Then

$$\|x\|^2 \geq \sum_{j \in F} |\langle x, u_j \rangle|^2 \geq S - \epsilon > \|x\|^2,$$

a contradiction. Therefore we must have

$$\sum_{j \in J^*} |\langle x, u_j \rangle|^2 \leq \|x\|^2,$$

that is, J^* is an upper bound of C.

By Zorn's Lemma, there exists a maximal element J_0 of I. We show that $J_0 = I$. If not, there exists $u_{i0} \notin J_0$. Let $y = \sum_{j\in J_0} \langle x, u_j\rangle u_j$. Then $u_{i0} \perp u_j$ for all $j \in J_0$, so that $\langle x - y, u_{i0}\rangle = \langle x, u_{i0}\rangle$. Thus

$$\|x\|^2 = \|x-y\|^2 + \|y\|^2 \geq |\langle x, u_{i0}\rangle|^2 + \sum_{j\in J_0} |\langle x, u_j\rangle|^2,$$

a contradiction to the maximality of J_0. This shows $J_0 = I$. ☐

Corollary 2.2.12. *If $\{u_1, u_2, \dots\}$ is an orthonormal sequence in an inner product space, then for each $x \in X$, $\lim_{n\to\infty} \langle x, u_n\rangle = 0$.*

Proof. By Theorem (2.2.11), $\sum_{n=1}^{\infty} |\langle x, u_n\rangle|^2$ converges. Thus $\langle x, u_n\rangle \to 0$ as $n \to \infty$. ☐

Corollary 2.2.13. *If $\{u_i : i \in I\}$ is an orthonormal system, then for each x at most a countable number of the Fourier coefficients $\langle x, u_i\rangle$ are nonzero.*

Proof. Fix $x \in X$ and put $J_n := \{i \in I : |\langle x, u_i\rangle| > 1/n\}$. By the Bessel inequality,

$$\|x\|^2 \geq \sum_{j\in J_n} |\langle x, u_j\rangle|^2 \geq \sum_{j\in J_n} 1/n^2 = \frac{\#J_n}{n^2}.$$

Thus J_n is a finite set. Since

$$\{i : \langle x, u_i\rangle \neq 0\} = \bigcup_{n=1}^{\infty} J_n,$$

we see that this set must be countable, being a union of countably many finite sets. ☐

Definition 2.2.14 (Orthonormal Basis)**.** *Let X be an inner product space. An* **orthonormal basis** *for X is any maximal orthonormal set in X. That is, any orthonormal set which is not properly contained in another orthonormal set.*

Theorem 2.2.15. *Every nontrivial inner product space has an orthonormal basis.*

Proof. Let X be the inner product space. Since $X \neq \{0\}$, there exists a nonzero vector $x \in X$. Note that $\{x/\|x\|\}$ is orthonormal. Let F be the family of all orthonormal subsets of X, and partially order this family by inclusion. Let C be a totally ordered subset in F. Put $A^* := \bigcup_{A\in C} A$. Then, for all $A \in C$, we have $A \subseteq A^*$, so that A^* is an upper bound for C. We check that A^* is orthonormal.

Let $x, y \in A^*$ be such that $x \neq y$. Then there exist $A_1, A_2 \in C$ such that $x \in A_1$ and $y \in A_2$. Since C is totally ordered, either $A_1 \subseteq A_2$ or $A_2 \subseteq A_1$, say $A_1 \subseteq A_2$. Then $x, y \in A_2$. Since A_2 is orthonormal, $\langle x, y\rangle = 0$ and $\|x\| = \|y\| = 1$. Thus A^* is orthonormal.

By Zorn's Lemma, it follows that there exists a maximal orthonormal set. ☐

Theorem 2.2.16 (Orthonormal Basis Theorem)**.** *Let $\{u_\alpha\}_{\alpha\in A}$ be an orthonormal family in a Hilbert space X. The following properties are equivalent:*

(1) $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal basis for X;
(2) If $x \in X$ and $x \perp u_\alpha$ for all $\alpha \in A$, then $x = 0$;
(3) For each $x \in X$, $x = \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha$;
(4) For each $x, y \in X$, $\langle x, y\rangle = \sum_{\alpha\in A} \langle x, u_\alpha\rangle \overline{\langle y, u_\alpha\rangle}$;
(5) For each $x \in X$, $\|x\|^2 = \sum_{\alpha\in A} |\langle x, u_\alpha\rangle|^2$ (Parseval Identity).

Proof. First suppose that $\{u_\alpha\}_{\alpha\in A}$ is an orthonormal basis for X. By contradiction, let $x \neq 0$ be such that $x \perp u_\alpha$ for all $\alpha \in A$. Then we may adjoin $x/\|x\|$ to the family $\{u_\alpha\}_{\alpha\in A}$ to obtain a larger orthonormal basis. Thus the original family is not maximal and is therefore not a basis.

Next, assume that if $x \in X$ and $x \perp u_\alpha$ for all $\alpha \in A$, then $x = 0$. Put $y := \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha$. By Bessel's Inequality,

$$\|y\|^2 \leq \sum_{\alpha\in A} \| \langle x, u_\alpha\rangle u_\alpha\|^2 = \sum_{\alpha\in A} | \langle x, u_\alpha\rangle |^2 \leq \|x\|^2.$$

By the Pythagorean Law, the series defining y converges. We find for all u_β, $\beta \in A$,

$$\begin{aligned}\langle x - y, u_\beta\rangle &= \langle x, u_\beta\rangle - \left\langle \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha, u_\beta \right\rangle = \langle x, u_\beta\rangle - \sum_{\alpha\in A} \langle x, u_\alpha\rangle \langle u_\alpha, u_\beta\rangle \\ &= \langle x, u_\beta\rangle - \sum_{\alpha\in A} \langle x, u_\alpha\rangle \delta_{\alpha\beta} = \langle x, u_\beta\rangle - \langle x, u_\beta\rangle = 0.\end{aligned}$$

By the assumption, $x - y = 0$.

Now suppose that for every $x \in X$, $x = \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha$. Put

$$x := \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha, \quad y := \sum_{\alpha\in A} \langle y, u_\alpha\rangle u_\alpha.$$

Then

$$\begin{aligned}\langle x, y\rangle &= \left\langle \sum_{\alpha\in A} \langle x, u_\alpha\rangle u_\alpha, \sum_{\beta\in A} \langle y, u_\beta\rangle u_\beta \right\rangle = \sum_{\alpha\in A} \langle x, u_\alpha\rangle \left\langle u_\alpha, \sum_{\beta\in A} \langle y, u_\beta\rangle u_\beta \right\rangle \\ &= \sum_{\alpha\in A}\sum_{\beta\in A} \langle x, u_\alpha\rangle \langle y, u_\beta\rangle \langle u_\alpha, u_\beta\rangle = \sum_{\alpha\in A}\sum_{\beta\in A} \langle x, u_\alpha\rangle \overline{\langle y, u_\beta\rangle} \delta_{\alpha\beta} \\ &= \sum_{\alpha\in A} \langle x, u_\alpha\rangle \overline{\langle y, u_\alpha\rangle}.\end{aligned}$$

Assume now that for all $x, y \in X$, $\langle x, y\rangle = \sum_{\alpha\in A} \langle x, u_\alpha\rangle \overline{\langle y, u_\alpha\rangle}$. Let $y = x$. Then we see that

$$\|x\|^2 = \sum_{\alpha\in A} | \langle x, u_\alpha\rangle |^2.$$

Finally, assume that $\|x\|^2 = \sum_{\alpha\in A} |\langle x, u_\alpha\rangle|^2$. By contradiction, suppose that $\{u_\alpha\}_{\alpha\in A}$ is not an orthonormal basis for X. Then $\{u_\alpha\}_{\alpha\in A}$ is not an orthonormal set. We may adjoin a new element, x, to obtain a larger orthonormal set. But then

$$1 = \|x\|^2 \neq \sum_{\alpha\in A} | \langle x, u_\alpha\rangle |^2 = 0,$$

a contradiction. ∎

Example 2.2.17. *One orthonormal basis in ℓ^2 is obtained by defining $\{u_{nk}\}_{k=1}^{\infty}$ by $u_{nk} = \delta_{nk}$ for all $n \in \mathbb{N}$. To see that this is an orthonormal basis, suppose that $x \in \ell^2$ and $\langle x, u_n \rangle = 0$ for all $n \in \mathbb{N}$. Then $x_n = 0$ for all $n \in \mathbb{N}$, so that $x = 0$.*

Recall that we have defined the orthogonal projection of a Hilbert space X onto a closed subspace Y to be the mapping P such that for each $x \in X$, Px is the point of Y closest to x.

Theorem 2.2.18 (Orthogonal Projection Theorem)**.** *The orthogonal projection P of a Hilbert space X onto a closed subspace Y has the following properties:*

(1) P is well–defined, that is, Px exists and is unique in Y;
(2) P is surjective, that is, $P(X) = Y$;
(3) P is linear;
(4) If $Y \neq \{0\}$, then $\|P\| = 1$;
(5) $x - Px \perp Y$ for all x;
(6) P is Hermitian, that is, $\langle Px, w \rangle = \langle x, Pw \rangle$ for all x and w;
(7) If $\{u_\alpha\}_{\alpha \in A}$ is an orthonormal basis for Y, then $Px = \sum_{\alpha \in A} \langle x, u_\alpha \rangle u_\alpha$;
(8) P is idempotent, that is, $P^2 = P$;
(9) $Py = y$ for all $y \in Y$, that is, $P|_Y = I_Y$;
(10) $\|x\|^2 = \|Px\|^2 + \|x - Px\|^2$.

Theorem 2.2.19. *Let $\{v_n\}_{n=1}^{\infty}$ be a linearly independent sequence in an inner product space X. Define*

$$u_1 := \frac{v_1}{\|v_1\|}, \quad u_n := \frac{v_n - \sum\limits_{k=1}^{n-1} \langle v_n, u_k \rangle u_k}{\|v_n - \sum\limits_{k=1}^{n-1} \langle v_n, u_k \rangle u_k\|}, \ n = 2, 3, \ldots.$$

Then $\{u_n\}_{n=1}^{\infty}$ is an orthonormal sequence, and for each $n \in \mathbb{N}$, $\operatorname{span}\{u_k\}_{k=1}^{n} = \operatorname{span}\{v_k\}_{k=1}^{n}$.

Example 2.2.20. *A normed linear space is said to be* **separable** *if it contains a countable dense subset. If an inner product space is nonseparable, it cannot have a countable orthonormal basis.*

Consider the uncountable family of functions $u_\lambda(t) := e^{i\lambda t}$, where $t, \lambda \in \mathbb{R}$. This family of functions is linearly independent and is therefore a Hamel basis for some linear space X. We introduce an inner product in X by defining the inner product of two elements in the Hamel basis:

$$\langle u_\lambda, u_\sigma \rangle := \delta_{\lambda\sigma}.$$

This is the family that arises in the following integration:

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} u_\lambda(t)\overline{u_\sigma(t)}\ dt = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{i(\lambda-\sigma)t}\ dt.$$

If $\lambda = \sigma$, this calculation yields the result 1. If $\lambda \neq \sigma$, we get zero.

Example 2.2.21. *We consider the space $\mathcal{C}[-1,1]$ with the inner product*

$$\langle f, g \rangle := \int_{-1}^{1} f(t)g(t)\ dt.$$

Applying the Gram–Schmidt process to the standard monomial basis $\{1, t, t^2, t^3, \dots\}$ *yields the* **Legendre polynomials**. *The unnormalized polynomials may be written recursively:*

$$P_0(t) := 1, \quad P_2(t) := t, \quad P_n(t) := \frac{2n-1}{n} t P_{n-1}(t) - \frac{n-1}{n} P_{n-2}(t), \ n \geq 2.$$

The corresponding orthonormal system is $p_n := P_n/\|P_n\|$. *The completion of the space* $\mathcal{C}[-1,1]$ *with respect to the norm induced by the inner product is the space* $L^2[-1,1]$. *Thus every function* $f \in L^2[-1,1]$ *is represented in the* L^2-*sense by the series*

$$f := \sum_{n=0}^{\infty} \langle f, p_n \rangle p_n.$$

2.3. **Linear Functionals and Operators.** Recall that a *linear functional* on a linear space X is a mapping ϕ from X to $\mathbb{R}$ (or $\mathbb{C}$) such that for all $x, y \in X$ and $a, b \in \mathbb{R}$,

$$\phi(ax + by) = a\phi(x) + b\phi(y).$$

Recall also that if X has a norm, and if

$$\sup_{\|x\|=1} |\phi(x)| < \infty,$$

we say that ϕ is a *bounded linear functional*, and we denote by $\|\phi\|$ the above inequality.

Theorem 2.3.1 (Riesz Representation Theorem)**.** *Every continuous linear functional defined on a Hilbert space* X *is of the form* $\phi(x) = \langle x, v \rangle$ *for some* $v \in X$ *that is determined uniquely by the given functional.*

Proof. Let $\phi : X \to \mathbb{R}$ be a continuous linear functional. Note first that if $\ker X = X$, then $\phi(x) \equiv 0$ and thus $\phi(x) = \langle x, 0 \rangle$. Otherwise, let $u \neq 0 \in \ker(\phi)^{\perp}$. Without loss of generality, we may assume that $\phi(u) = 1$. Observe that $X = \ker(\phi) \oplus \mathbb{R}u$, for we have that

$$x = x - \phi(x)u + \phi(x)u,$$

and $x - \phi(x)u \in \ker(\phi)$. Put $v := u/\|u\|^2$. Then

$$\begin{aligned} \langle x, v \rangle &= \langle x - \phi(x)u + \phi(x)u, v \rangle = \langle x - \phi(x)u, v \rangle + \langle \phi(x)u, v \rangle \\ &= \phi(x) \langle u, v \rangle = \phi(x) \left\langle u, \frac{u}{\|u\|^2} \right\rangle = \phi(x) \langle u, u \rangle / \|u\|^2 = \phi(x). \end{aligned}$$

This completes the proof. □

Example 2.3.2. *Let* X *be a finite–dimensional Hilbert space with a basis* $\{u_1, u_2, \dots, u_n\}$, *not necessarily orthonormal. Each point* $x \in X$ *can be represented uniquely in the form*

$$x = \sum_{j=1}^{n} \lambda_j(x) u_j,$$

where λ_j, $j = 1, 2, \dots, n$ *are continuous linear functionals. Hence by Theorem* (2.3.1) *there exist points* $v_j \in X$ *such that*

$$x = \sum_{j=1}^{n} \langle x, v_j \rangle u_j,$$

for all $x \in X$.

Moreover, since $u_i = \sum_{j=1}^{n} \langle u_i, v_j \rangle u_j$*, we must have* $\langle u_i, v_j \rangle = \delta_{ij}$*. In this situation, we say that the two sets* $\{u_1, \ldots, u_n\}$ *and* $\{v_1, \ldots, v_n\}$ *are mutually* **biorthogonal** *or that they form a biorthogonal pair.*

Example 2.3.3. *The orthogonal projection* P *of a Hilbert space* X *onto a closed subspace* Y *is a bounded linear operator from* X *into* X*. Theorem* (2.2.6) *shows that* P *has several nice properties.*

Example 2.3.4. *It is easy to create bounded linear operators on a Hilbert space* X*. Take any orthonormal system* $\{u_j\}_{j\in I}$*, possibly finite, countable, or uncountable. Define*

$$Ax := \sum_{i\in I} \sum_{j\in I} a_{ij} \langle x, u_j \rangle u_i.$$

If the coefficients a_{ij} *have the property* $\sum_{i\in I} \sum_{j\in I} |a_{ij}|^2 < \infty$*, then* A *will be continuous.*

Theorem 2.3.5 (Existence of Adjoints)**.** *If* A *is a bounded linear operator on a Hilbert space* X*, then there is a uniquely defined bounded linear operator* A^* *such that*

$$\langle Ax, y \rangle = \langle x, A^* y \rangle$$

for all $x, y \in X$*. Moreover,* $\|A^*\| = \|A\|$*.*

Proof. Fix $y \in X$, and notice that the mapping $\phi : X \to \mathbb{R}$ defined by $\phi(x) = \langle Ax, y \rangle$ is a bounded linear functional on X :

$$\begin{aligned} \langle A(\lambda x + \mu z), y \rangle &= \langle \lambda Ax + \mu Az, y \rangle = \lambda \langle Ax, y \rangle + \mu \langle Az, y \rangle \\ &\le \lambda \|Ax\| \|y\| + \mu \|Az\| \|y\| \le \|A\| \|y\| (\lambda \|x\| + \mu \|z\|). \end{aligned}$$

Thus, by the Riesz Representation Theorem (2.3.1), there exists a unique vector v such that $\langle Ax, y \rangle = \langle x, v \rangle$ for all $x \in X$. Noting that v depends on A and y, we denote $v =: A^* y$. We show that the mapping A^* is linear and bounded.

We first show linearity. By the Lemma (2.1.8), it suffices to show that for all $x \in X$,

$$\langle x, A^*(\lambda y + \mu z) \rangle = \langle x, \lambda A^* y + \mu A^* z \rangle .$$

We see that, by definition of A^*, ,

$$\begin{aligned} \langle x, A^*(\lambda y + \mu z) \rangle &= \langle Ax, \lambda y + \mu z \rangle = \langle Ax, \lambda y \rangle + \langle Ax, \mu z \rangle \\ &= \overline{\lambda} \langle Ax, y \rangle + \overline{\mu} \langle Ax, z \rangle = \overline{\lambda} \langle x, A^* y \rangle + \overline{\mu} \langle x, A^* z \rangle \\ &= \langle x, \lambda A^* y + \mu A^* z \rangle , \end{aligned}$$

which shows linearity.

We now show boundedness. By the Lemma (2.1.8), we have

$$\begin{aligned} \|A^*\| &= \sup_{\|y\|=1} \|A^* y\| = \sup_{\|y\|=1} \sup_{\|x\|=1} | \langle x, A^* y \rangle | \\ &= \sup_{\|x\|=1} \sup_{\|y\|=1} | \langle x, A^* y \rangle | \\ &= \sup_{\|x\|=1} \sup_{\|y\|=1} | \langle Ax, y \rangle | \\ &\le \sup_{\|x\|=1} \sup_{\|y\|=1} \|Ax\| \|y\| \end{aligned}$$

$$= \sup_{\|x\|=1} \|Ax\| = \|A\|.$$

Finally, to show uniqueness, suppose that there exists $B \in \mathcal{L}(X, X)$ such that

$$\langle Ax, y \rangle = \langle x, By \rangle$$

for all $x, y \in X$. Then, for any $x \in X$, it follows

$$\langle x, A^* y \rangle = \langle Ax, y \rangle = \langle x, By \rangle$$

for all $y \in X$. Thus by the Lemma (2.1.8), we have $A^* \equiv B$.

This completes the proof. □

Definition 2.3.6 (Adjoint). *Let X be a Hilbert space and let $A \in \mathcal{L}(X, X)$. Then the operator A^* such that*

$$\langle Ax, y \rangle = \langle x, A * *y \rangle$$

for all $x, y \in X$ is called the **adjoint** *of A.*

Recall that for an operator A on a Banach space X, A^* is defined on X^* by the equation

$$A^* \phi = \phi \circ A.$$

If X is a Hilbert space, then X^* may be identified with X by the Riesz Representation Theorem, that is, for any $\phi \in X^*$, $\phi(x) = \langle x, y \rangle$ for some unique $y \in X$. Thus

$$(A^* \phi)(x) = (\phi \circ A)(x) = \phi(Ax) = \langle Ax, y \rangle .$$

Hence, the Hilbert space adjoint coincides with the Banach space definition.

Example 2.3.7. *Let S be any measure space, and let an operator T on $L^2(S)$ be defined by the equation*

$$(Tx)(s) := \int_S k(s,t) x(t)\ dt.$$

Assume that the kernel k is $L^2(S) \times L^2(S)$ integrable, in the sense that

$$\int_S \int_S |k(s,t)|^2\ dt\ ds < \infty.$$

Then T is bounded, and its adjoint T^* is an integral operator of the same type. More specifically,

$$(T^* x)(s) = \int_S \overline{k(s,t)} x(t)\ dt.$$

Definition 2.3.8 (Self–Adjoint Operator). *If A is a bounded linear operator such that $A = A^*$, then we say that A is* **self–adjoint**.

Definition 2.3.9 (Hermitian Operator). *A linear operator on an inner–product space is said to be* **Hermitian** *if $\langle Ax, y \rangle = \langle x, Ay \rangle$ for all x and y.*

Note that by definition, a Hermitian operator is not necessarily bounded.

Theorem 2.3.10. *Let X be a Hilbert space. If A is a Hermitian operator on X, that is, if A is a linear map such that $\langle Ax, y \rangle = \langle y, Ax \rangle$ for all $x, y \in X$, then A is bounded and self–adjoint.*

Proof. For every fixed $y \in X$ such that $\|y\| \leq 1$, define a functional ϕ_y by $\phi_y(x) := \langle Ax, y \rangle$. By the linearity of A, ϕ_y is obviously linear, and by the Cauchy–Schwarz Inequality, we have

$$|\phi_y(x)| = |\langle Ax, y \rangle| = |\langle x, Ay \rangle| \leq \|x\|\|Ay\| < \infty.$$

Thus ϕ_y is bounded. Moreover, by the Lemma (2.1.8), we see that

$$\sup_{\|y\|=1} \|\phi_y(x)\| = \sup_{\|y\|\leq 1} |\langle Ax, y \rangle| = \|Ax\|.$$

Thus, by the Uniform Boundedness Principle, $\sup_{\|y\|\leq 1} \|\phi_y\| < \infty$, and thus

$$\begin{aligned} \infty > \sup_{\|y\|\leq 1} \|\phi_y\| &= \sup_{\|y\|\leq 1} \sup_{\|x\|\leq 1} |\phi_y(x)| \\ &= \sup_{\|y\|\leq 1} \sup_{\|x\|\leq 1} |\langle Ax, y \rangle| \\ &= \sup_{\|x\|\leq 1} \sup_{\|y\|\leq 1} |\langle Ax, y \rangle| \\ &= \sup_{\|x\|\leq 1} \|Ax\| = \|A\|. \end{aligned}$$

Thus, A is bounded.

Next, the equation

$$\langle x, A^*y \rangle = \langle Ax, y \rangle = \langle x, Ay \rangle,$$

along with uniqueness of the adjoint, shows that $A \equiv A^*$. This completes the proof. □

Definition 2.3.11. *Let X be a Hilbert space and let A be a bounded linear operator on X. We define a new norm $|||\cdot|||$ by*

$$|||A||| := \sup_{\|x\|=1} |\langle Ax, x \rangle|.$$

Lemma 2.3.12 (Generalized Cauchy–Schwarz Inequality)**.** *If A is a Hermitian operator on a Hilbert space X, then, for any $x, y \in X$,*

$$|\langle Ax, y \rangle| \leq |||A|||\|x\|\|y\|.$$

Proof. First observe that

$$\langle A(x+y), x+y \rangle = \langle Ax, x \rangle + \langle Ax, y \rangle + \langle Ay, x \rangle + \langle Ay, y \rangle \tag{2.3.0.1}$$

$$-\langle A(x-y), x-y \rangle = -\langle Ax, x \rangle + \langle Ax, y \rangle + \langle Ay, x \rangle - \langle Ay, y \rangle. \tag{2.3.0.2}$$

Adding equations (2.3.0.1) and applying the Hermitian property of A gives

$$\langle A(x+y), x+y \rangle - \langle A(x-y), x-y \rangle = 4\mathrm{Re}\,\langle Ax, y \rangle.$$

Thus we find, for $x \neq 0$,

$$|\langle Ax, x \rangle| = \|x\|^2 \left\langle A\left(\frac{x}{\|x\|}\right), \frac{x}{\|x\|} \right\rangle \leq |||A|||\|x\|^2.$$

By the above and the parallelogram law we have

$$\begin{aligned} |4\mathrm{Re}\,\langle Ax, y \rangle| &= |\langle A(x+y), x+y \rangle - \langle A(x-y), x-y \rangle| \\ &\leq |\langle A(x+y), x+y \rangle| + |\langle A(x-y), x-y \rangle| \\ &\leq |||A|||\|x+y\|^2 + |||A|||\|x-y\|^2 \\ &= |||A|||(2\|x\|^2 + 2\|y\|^2). \end{aligned}$$

Letting $\|x\| = \|y\| = 1$ in the preceding argument shows that

$$|\mathrm{Re}\,\langle Ax, y\rangle\,| \leq |||A|||.$$

Finally, fix $x, y \in X$. Note that we may choose $\theta \in \mathbb{C}$ such that $|\theta| = 1$ and $\theta\,\langle Ax, y\rangle = |\,\langle Ax, y\rangle\,|$. Then

$$|\,\langle Ax, y\rangle\,| = \theta\,\langle Ax, y\rangle = \langle A(\theta x), y\rangle = |\mathrm{Re}\,\langle A(\theta x), y\rangle\,| \leq |||A|||.$$

By homogeneity, this completes the proof. ☐

Lemma 2.3.13. *If A is Hermitian, then $\|A\| = |||A|||$.*

Proof. By the Cauchy–Schwarz Inequality, we first note that

$$|||A||| = \sup_{\|u\|=1} |\,\langle Au, u\rangle\,| \leq \sup_{\|u\|=1} \|Au\|\|u\| = \sup_{\|u\|=1} \|Au\| = \|A\|.$$

For the reverse inequality, we apply the preceding Lemma (2.3.12) and observe that

$$\begin{aligned}\|A\| &= \sup_{\|x\|=1} \|Ax\| = \sup_{\|x\|=1}\sup_{\|y\|=1} |\,\langle Ax, y\rangle\,| \\ &\leq \sup_{\|x\|=1}\sup_{\|y\|=1} |||A|||\|x\|\|y\| = |||A|||.\end{aligned}$$

☐

Definition 2.3.14 (Compact Operator)**.** *Let X and Y be normed linear spaces, and let $\Sigma \subseteq X$ denote the unit ball. An operator $A : X \to Y$ is said to be* **compact** *if the set $A(\Sigma) \subseteq Y$ is relatively compact in Y ($\overline{A(\Sigma)}$ is compact).*

Recall that a continuous linear operator is an operator that maps the unit ball in the domain to a bounded set in the codomain. Thus, compactness of an operator is a stronger condition than continuity.

Lemma 2.3.15. *Every continuous linear operator from a normed linear space to a finite–dimensional normed linear space is compact.*

Proof. Let X be a normed linear space, and let A be a continuous linear operator from X to a finite–dimensional normed linear space Y. Let $\Sigma \subseteq X$ be the unit ball. Since A is continuous, $A(\Sigma)$ is a bounded set in Y. Thus, $\overline{A(\Sigma)}$ is a closed and bounded set in the finite–dimensional space Y, and hence A is compact. This completes the proof. ☐

Theorem 2.3.16. *Let X and Y be Banach spaces. Then the set of compact operators in $\mathcal{L}(X, Y)$ is closed.*

Proof. Let $\{A_n\}_{n=1}^\infty$ be a sequence of compact operators from X to Y, and suppose that there exists $A \in \mathcal{L}(X, Y)$ such that $\lim_{n\to\infty} \|A_n - A\| = 0$.

We show that A is compact. Denote by $\Sigma \subseteq X$ the unit ball, and let $\{x_n\}_{n=1}^\infty$ be a sequence in Σ. It suffices to find a convergent subsequence in $\{Ax_n\}_{n=1}^\infty$.

Since A_1 is compact, there exists an increasing subsequence $I_1 \subseteq \mathbb{N}$ such that $\{A_1 x_n\}_{n\in I_1}$ converges. Likewise, there is an increasing subsequence $I_2 \subseteq I_1$ such that the sequence $\{A_2 x_n\}_{n\in I_2}$ converges. Continue in this fashion. Applying Cantor's diagonalization argument, we let $I \subseteq \mathbb{N}$ be the sequence whose $n-$th term is the $n-$th term of I_n, $n \in \mathbb{N}$. By this construction, the sequence $\{A_i x_n\}_{n\in I}$ converges for all $i \in \mathbb{N}$.

Finally, we show that $\{Ax_n\}_{n\in I}$ converges. It suffices to show that $\{Ax_n\}_{n\in I}$ is Cauchy. We find that

$$\begin{aligned}\|Ax_n - Ax_m\| &\le \|Ax_n - A_ix_n\| + \|A_ix_n - A_ix_m\| + \|A_ix_m - Ax_m\| \\ &\le \|A - A_i\|\|x_n\| + \|A_ix_n - A_ix_m\| + \|A_i - A\|\|x_m\|.\end{aligned}$$

This completes the proof. $\square$

Theorem 2.3.17. *Let S be any measure space. Define the operator $T : L^2(S) \to L^2(S)$ by*

$$(Tx)(s) := \int_S k(s,t)x(t)\ dt.$$

If the kernel k is in $L^2(S\times S)$, then T is compact.

Proof. Choose an orthonormal basis $\{u_n\}_{n=1}^{\infty}$ for $L^2(S)$, and define $a_{nm} := \langle Tu_m, u_n\rangle$. Note, for any $x \in L^2(S)$, $x = \sum_{n=1}^{\infty} \langle x, u_n\rangle u_n$, it follows

$$\begin{aligned}Tx = \sum_{n=1}^{\infty} \langle Tx, u_n\rangle u_n &= \sum_{n=1}^{\infty} \left\langle \sum_{m=1}^{\infty} \langle x, u_m\rangle Tu_m, u_n \right\rangle u_n \\ &= \sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} a_{nm} \langle x, u_m\rangle \right) u_n.\end{aligned}$$

Observe that

$$\begin{aligned}\|k\|^2 &= \int\int |k(s,t)|^2\ dtds = \int \|k(s,\cdot)\|^2\ ds = \int \sum_{n=1}^{\infty} |\langle k(s,\cdot), u_n\rangle|^2\ ds \\ &= \int \sum_{n=1}^{\infty} \left|\int k(s,t)u_n(t)\ dt\right|^2\ ds = \int \sum_{n=1}^{\infty} |(Tu_n)(s)|^2\ ds \\ &= \sum_{n=1}^{\infty} \int |(Tu_n)(s)|^2\ ds = \sum_{n=1}^{\infty} \|Tu_n\|^2 \\ &= \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |\langle Tu_n, u_m\rangle|^2 = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |a_{mn}|^2 =: \sum_{m=1}^{\infty} \beta_m,\end{aligned}$$

where $\beta_m := \sum_{n=1}^{\infty} |a_{mn}|^2$. We truncate the series that defines T in order to obtain operators with finite–dimensional range that approximate T. Thus, we put

$$T_nx := \sum_{i=1}^{n}\sum_{j=1}^{\infty} a_{ij} \langle x, u_j\rangle u_i.$$

By subtraction,

$$Tx - T_nx = \sum_{i>n}\sum_{j=1}^{\infty} a_{ij} \langle x, u_j\rangle u_i.$$

Further, by the Cauchy–Schwarz Inequality and the Bessel Inequality, we have

$$\begin{aligned}\|Tx - T_nx\|^2 = \sum_{i>n} \left|\sum_{j=1}^{\infty} a_{ij} \langle x, u_j\rangle\right|^2 &\le \sum_{i>n}\sum_{j=1}^{\infty} |a_{ij}|^2 \sum_{k=1}^{\infty} |\langle x, u_k\rangle|^2 \\ &\le \|x\|^2 \sum_{i>n}\sum_{j=1}^{\infty} |a_{ij}|^2 = \|x\|^2 \sum_{i>n} \beta_i.\end{aligned}$$

This shows that $\lim_{n\to\infty} \|T - T_n\| = 0$. Since each T_n has finite–dimensional range, each T_n is compact, and thus T is compact by Theorem (2.3.16). $\square$

Theorem 2.3.18. *Let X be a Hilbert space and let A be a bounded linear operator on X. Then $\mathcal{N}(A) = \mathcal{R}(A^*)^\perp$.*

Proof. Let $x \in \mathcal{N}(A)$ and let $z \in X$ be arbitrary. We find

$$\langle x, A^* z\rangle = \langle Ax, z\rangle = \langle 0, z\rangle = 0.$$

Thus $x \in \mathcal{R}(A^*)^\perp$ and $\mathcal{N}(A) \subseteq \mathcal{R}(A^*)^\perp$.

On the other hand, if $x \in \mathcal{R}(A^*)^\perp$, then

$$\langle Ax, Ax\rangle = \langle x, A^*(Ax)\rangle = 0.$$

Thus $Ax = 0$, $x \in \mathcal{N}(A)$, and $\mathcal{R}(A^*)^\perp \subseteq \mathcal{N}(A)$, which completes the proof. $\square$

Corollary 2.3.19. *A Hermitian operator whose range is dense is injective.*

Definition 2.3.20 (Weak Convergence (Hilbert Space))**.** *A sequence $\{x_n\}_{n=1}^\infty$ in a Hilbert space X is said to* **converge weakly** *to a point $x \in X$ if for all $y \in X$,*

$$\lim_{n\to\infty} \langle x_n, y\rangle = \langle x, y\rangle .$$

We write $x_n \rightharpoonup x$.

Example 2.3.21. *If $\{u_n\}_{n=1}^\infty$ is an orthonormal sequence in a Hilbert space, then $u_n \rightharpoonup 0$. This follows from Bessel's Inequality,*

$$\sum_{n=1}^\infty |\langle u_n, y\rangle|^2 \le \|y\|^2,$$

which shows that $\lim_{n\to\infty} \langle u_n, y\rangle = 0$ for all y.

Definition 2.3.22 (Weakly Cauchy (Hilbert Space))**.** *A sequence $\{x_n\}_{n=1}^\infty$ in a Hilbert space X is said to be* **weakly Cauchy** *if, for all $y \in X$, the sequence $\{\langle x_n, y\rangle\}_{n=1}^\infty$ is Cauchy in $\mathbb{C}$.*

Lemma 2.3.23. *A weakly Cauchy sequence in a Hilbert space is weakly convergent to a point in the Hilbert space.*

Proof. Let X be a Hilbert space, and let $\{x_n\}_{n=1}^\infty$ be weakly Cauchy in X. For each $y \in X$, the sequence $\{\langle y, x_n\rangle\}_{n=1}^\infty$ has the Cauchy property in $\mathbb{C}$, and is therefore bounded in $\mathbb{C}$. The linear functionals $\phi_n : X \to \mathbb{C}$ defined by $\phi_n(y) := \langle y, x_n\rangle$ consequently have the property

$$\sup_{n\in\mathbb{N}} |\phi_n(y)| < \infty,$$

for all $y \in X$. By the Uniform Boundedness Principle, we have that $\|\phi_n\| \le M$ for some $M \in \mathbb{R}$. Since

$$\|x_n\| = \sup_{\|y\|=1} |\langle y, x_n\rangle| = \|\phi_n\| \le M,$$

we conclude that $\{x_n\}_{n=1}^\infty$ is a bounded sequence. Put $\phi(y) := \lim_{n\to\infty} \langle y, x_n\rangle$. Then ϕ is a bounded linear functional on X. By the Riesz Representation Theorem, there exists $x \in X$ such that $\phi(y) = \langle y, x\rangle$. Hence, $\lim_{n\to\infty} \langle y, x_n\rangle = \langle y, x\rangle$ and thus $x_n \rightharpoonup x$. $\square$

Theorem 2.3.24 (Fredholm Alternative, Hilbert Space)**.** *Let A be a continuous linear operator on a Hilbert space. If the range of A is closed, then it is the orthogonal complement of the null space of A^*. That is,*

$$\mathcal{R}(A) = \mathcal{N}(A^*)^{\perp}.$$

Proof. Let $x \in \mathcal{R}(A)$. Then $x = Au$ for some $u \in X$. Thus, for all $z \in \mathcal{N}(A^*)$, we have

$$\langle x, y \rangle = \langle Au, y \rangle = \langle u, A**y \rangle = 0.$$

For the reverse inclusion, we proceed by contrapositive and suppose that $x \notin \mathcal{R}(A)$. Since $\mathcal{R}(A)$ is a closed subspace by the supposition, there exists a continuous linear functional ϕ such that $\phi(x) = 1$ and $\phi(z) = 0$ for all $z \in \mathcal{R}(A)$. By the Riesz Representation Theorem, there exists $y \in X$ such that $\phi(u) = \langle u, y \rangle$ for all $u \in X$. Note that for any $z \in \mathcal{R}(A)$, we may write $z = Av$ for some $v \in X$, and thus

$$\phi(z) = \langle z, y \rangle = \langle Av, y \rangle = \langle v, A**y \rangle = 0,$$

so that $A^*y = 0$ and $y \in \mathcal{N}(A^*)$. But since $\langle x, y \rangle \neq 0$, we have $x \notin \mathcal{N}(A^*)^{\perp}$, which completes the proof. ☐

2.4. **Spectral Theory.** In this section we study the structure of linear operators on a Hilbert space. Ideally, we want to dissect an operator into a sum of simple operators or a series of simple operators. Specifically, we consider operators of the form

$$Lx = \sum_{j=1}^{\infty} a_j \langle x, u_j \rangle u_j.$$

Definition 2.4.1 (Eigenvalue)**.** *An* **eigenvalue** *of an operator A is a complex number λ such that $A - \lambda I$ has a nontrivial null space.*

We denote by $\Lambda(A)$ the set of all eigenvalues of an operator A.

If X is a finite–dimensional space, and if $A : X \to X$ is a linear operator, then A certainly has eigenvalues. To see this, introduce a basis $\{u_n\}_{n=1}^N$ for X so that A can be identified with a square matrix. Then the following conditions on a complex number λ are equivalent:

(1) $A - \lambda I$ has a nontrivial null space;
(2) $A - \lambda I$ is singular;
(3) $\det(A - \lambda I) = 0$.

Note that an operator on an infinite–dimensional space may have no eigenvalues.

Definition 2.4.2 (Eigenvector)**.** *Let λ be an eigenvalue of an operator A. Then any nontrivial solution of the equation $Ax = \lambda x$ is called an* **eigenvector** *of A belonging to the eigenvalue λ.*

Lemma 2.4.3. *If A is a Hermitian operator on an inner–product space, then:*

(1) All eigenvalues of A are real;
(2) Any two eigenvectors of A associated with distinct eigenvalues of A are orthogonal to each other;
(3) The quadratic form $x \mapsto \langle Ax, x \rangle$ is real–valued.

Proof. Let $Ax = \lambda x$, $Ay = \mu y$, with $x, y \neq 0$, $\lambda \neq \mu$. Then

$$\lambda \langle x, x\rangle = \langle \lambda x, x\rangle = \langle Ax, x\rangle = \langle x, Ax\rangle = \langle x, \lambda x\rangle = \overline{\lambda} \langle x, x\rangle .$$

Thus $\lambda \in \mathbb{R}$. To see that $\langle x, y\rangle = 0$, we find that

$$(\lambda - \mu) \langle x, y\rangle = \langle \lambda x, y\rangle - \langle x, \mu y\rangle = \langle Ax, y\rangle - \langle x, Ay\rangle = 0.$$

Finally, note that

$$\langle Ax, x\rangle = \overline{x, Ax} = \overline{Ax, x}.$$

□

Lemma 2.4.4. *A compact Hermitian operator A on an inner product space has at least one eigenvalue λ such that $|\lambda| = \|A\|$.*

Proof. The case $A = 0$ is trivial. Thus we assume that $A \neq 0$.

Recall from Lemma (2.3.13) that $|||A||| = \|A\|$. By definition of the supremum, there exists a sequence of points $\{x_n\}_{n=1}^{\infty}$ such that $\|x_n\| = 1$ and $\lim_{n\to\infty} |\langle Ax_n, x_n\rangle| = |||A|||$. Since A is compact, there exists a subsequence $\{x_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty} Ax_{n_k}$ exists. Put $y := \lim_{k\to\infty} Ax_{n_k}$. Note $y \neq 0$ because $|\langle Ax_{n_k}, x_{n_k}\rangle| \to |||A||| \neq 0$. By taking a further subsequence by the Bolzano–Weierstrass Theorem if necessary, we may assume that the limit $\lambda = \lim_{k\to\infty} Ax_{n_k}, x_{n_k}$ exists. By Lemma (2.4.3), λ is real. Then

$$\|Ax_{n_k} - \lambda x_{n_k}\|^2 = \|Ax_{n_k}\|^2 - \lambda \langle Ax_{n_k}, x_{n_k}\rangle - \lambda \langle x_{n_k}, Ax_{n_k}\rangle + \lambda^2 \|x_{n_k}\|^2.$$

Hence

$$0 \leq \lim_{k\to\infty} \|Ax_{n_k} - \lambda x_{n_k}\|^2 = \|y\|^2 - \lambda^2 - \lambda^2 + \lambda^2 = \|y\|^2 - \lambda^2.$$

Taking square roots, we find $|\lambda| \leq \|y\|$. For the reverse inequality, from the above argument we also find

$$\|y\| = \lim_{k\to\infty} \|Ax_{n_k}\| \leq \lim_{k\to\infty} \|A\|\|x_{n_k}\| = \|A\| = |\lambda|.$$

Thus we have $0 \leq \lim_{k\to\infty} \|Ax_{n_k} - \lambda x_{n_k}\| \leq 0$, and that

$$\|y - \lambda x_{n_k}\| \leq \|y - Ax_{n_k}\| + \|Ax_{n_k} - \lambda x_{n_k}\| \to 0.$$

Thus $x_{n_k} \to y/\lambda$. Finally, $Ay = A(\lim_{k\to\infty} \lambda x_{n_k}) = \lambda \lim_{k\to\infty} Ax_{n_k} = \lambda y$, so that y is in the null space of $A - \lambda I$. Hence λ is an eigenvalue of A. □

Theorem 2.4.5 (Spectral Theorem)**.** *If A is a compact Hermitian operator defined on an inner product space, then A is of the form $Ax = \sum_k \lambda_k \langle x, e_k\rangle e_k$ for an appropriate orthonormal sequence $\{e_k\}$ (possibly finite) and appropriate real numbers λ_k satisfying $\lim_{k\to\infty} \lambda_k = 0$. Moreover, the equations $Ae_k = \lambda_k e_k$ hold.*

Proof. If $A = 0$, the conclusion is trivial. Thus we assume that $A \neq 0$.

Let $X_1 := X$. Let λ_1 and e_1 be an eigenvalue and corresponding eigenvector determined by the preceding Lemma (2.4.4). Note that $|\lambda_1| = \|A\|$. Let $X_2 := \{x \in X : \langle x, e_1\rangle = 0\}$. Then X_2 is a subspace of X_1, and A maps X_2 into itself, for

$$\langle Ax, e_1\rangle = \langle x, Ae_1\rangle = \langle x, \lambda_1 e_1\rangle = \overline{\lambda_1} \langle x, e_1\rangle = 0$$

for any $x \in X_2$. We consider the restriction of A to the inner product space X_2, denoted by $A|_{X_2}$. This operator is also compact and Hermitian. Also, $\|A|_{X_2}\| \leq \|A\|$. If $A|_{X_2} \neq 0$, then the preceding lemma produces λ_2 and e_2, where $\|e_2\| = 1$, $|\lambda_2| = \|A|_{X_2}\| \leq |\lambda_1|$, $e_2 \perp X_1$,

$Ae_2 = \lambda_2 e_2$. Continue in this fashion. At the $n-$th stage we obtain $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| > 0$, $\{e_1, e_2, \ldots, e_n\}$ orthonormal, and $Ae_k = \lambda_k e_k$ for $k = 1, 2, \ldots, n$. We define X_{n+1} to be the orthogonal complement of the linear span of $\{e_1, \ldots, e_n\}$. If $A|_{X_{n+1}} = 0$, the process stops. Then the range of A is spanned by $e_1, \ldots, e_n$. That is, for any $x \in X$, the vector $x - \sum_{k=1}^n \langle x, e_k \rangle e_k$ is orthogonal to $\{e_1, \ldots, e_n\}$. Thus, it lies in X_{n+1}, and so A maps it to zero. That is,

$$Ax = \sum_{k=1}^{n} \langle x, e_k \rangle Ae_k = \sum_{k=1}^{n} \lambda_k \langle x, e_k \rangle e_k.$$

If $A|_{X_{n+1}} \neq 0$, we apply the preceding lemma to get λ_{n+1} and e_{n+1}.

It remains to show that if the above process does not terminate, then $\lim_{k\to\infty} \lambda_k = 0$. by contradiction, suppose that $|\lambda_n| \geq \epsilon > 0$ for all $n \in \mathbb{N}$. Then $\{e_n/\lambda_n\}_{n=1}^\infty$ is a bounded sequence, and by the compactness of A, the sequence $\{A(e_n/\lambda_n)\}_{n=1}^\infty$ must contain a convergent subsequence. But this is impossible, since $A(e_n/\lambda_n) = e_n$, and, $\{e_n\}_{n=1}^\infty$, being an orthonormal sequence, satisfies $\|e_n - e_m\| = \sqrt{2}$. In the infinite case, let $y_n := x - \sum_{k=1}^n \langle x, e_k \rangle e_k$. Since $y_n \perp \sum_{k=1}^n \langle x, e_k \rangle e_k$,

$$\|x\|^2 = \left\| y_n + \sum_{k=1}^{n} \langle x, e_k \rangle e_k \right\|^2 = \|y_n\|^2 + \sum_{k=1}^{n} |\langle x, e_k \rangle|^2 \geq \|y_n\|^2.$$

Since $|\lambda_{n+1}|$ is the norm of $\|A|_{X_{n+1}}\|$, we have

$$\|Ay_n\| \leq \|A|_{X_{n+1}}\| \|y_n\| \leq |\lambda_{n+1}| \|x\| \to 0.$$

Since $Ay_n = Ax - \sum_{k=1}^n \lambda_k \langle x, e_k \rangle e_k$, we have

$$Ax = \lim_{n\to\infty} \sum_{k=1}^{n} \lambda_k \langle x, e_k \rangle e_k.$$

This completes the proof. □

Corollary 2.4.6. *Every nonzero eigenvalue of A is in the sequence $\{\lambda_n\}_n$*

Proof. By contradiction, suppose that $Ax = \lambda x$, $x \neq 0$, $\lambda \neq 0$, and $\lambda \notin \{\lambda_n\}$. Then $x \perp e_n$ for all n by Lemma (2.4.3). But then $Ax = \sum_n \lambda_n \langle x, e_n \rangle e_n = 0$, a contradiction. □

Corollary 2.4.7. *Every nonzero eigenvalue λ of A occurs in the sequence $\{\lambda_n\}$ repeated a number of times equal to* $\dim\{x \in X : (A - \lambda I)x = 0\}$. *Each of these numbers is finite.*

Proof. Since $\lim_{n\to\infty} \lambda_n = 0$, a nonzero eigenvalue λ can be repeated only a finite number of times in the sequence. If it is repeated N times, then the subspace $\{x \in X : (A - \lambda I)x = 0\}$ contains an orthonormal set of N elements and so has dimension at least N. If the dimension were greater than N, then there would exist $x \neq 0$ such that $Ax = \lambda x$ and $\langle x, e_n \rangle = 0$ for all n, which is impossible. □

The following theorem provides an application of the spectral resolution of an operator, specifically, a formula for inverting the operator $A - \lambda I$ when A is compact and λ is not an eigenvalue of A.

Theorem 2.4.8. *Let A be a compact operator on an inner product space X having spectral decomposition $Ax = \sum_n \lambda_n \langle x, e_n \rangle e_n$. If $0 \neq \lambda \notin \{\lambda_n\}$, then $A - \lambda I$ is invertible, and*

$$(A - \lambda I)^{-1} x = -\frac{1}{\lambda} x + \frac{1}{\lambda} \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n.$$

Proof. Define the operator B by

$$Bx := -\frac{1}{\lambda} x + \frac{1}{\lambda} \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n.$$

Note that if the series defining B converges, then the conclusion follows, for we see that

$$\begin{aligned}
(A - \lambda I) Bx &= (A - \lambda I) \left\{ -\frac{1}{\lambda} x + \frac{1}{\lambda} \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n \right\} \\
&= -\frac{1}{\lambda} Ax + \frac{1}{\lambda} \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} Ae_n + x - \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n \\
&= x - \frac{1}{\lambda} \sum_n \frac{\lambda_n (\lambda_n - \lambda)}{\lambda_n - \lambda} \langle x, e_n \rangle e_n + \frac{1}{\lambda} \sum_n \lambda_n^2 \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n - \sum_n \lambda_n \frac{\langle x, e_n \rangle}{\lambda_n - \lambda} e_n \\
&= x.
\end{aligned}$$

In order to see that the series converges, define the partial sums

$$v_n := \sum_{k=1}^{n} \frac{\langle x, e_k \rangle}{\lambda_k - \lambda} e_k.$$

The sequence $\{v_n\}_{n=1}^{\infty}$ is bounded, since, by the Pythagorean Law and Bessel's Inequality we have

$$\|v_n\|^2 = \sum_{k=1}^{n} \left| \frac{\langle x, e_k \rangle}{\lambda_k - \lambda} \right|^2 \leq \sup_{j \in \mathbb{N}} \left| \frac{1}{\lambda_j - \lambda} \right|^2 \sum_{k=1}^{\infty} | \langle x, e_k \rangle |^2 \leq \beta \|x\|^2,$$

where

$$\beta := \sup_{j \in N} \left| \frac{1}{\lambda_j - \lambda} \right|^2.$$

Since A is compact, $\{\lambda_n\}$ converges to zero. Thus $\beta < \infty$. Moreover, the sequence $\{Av_n\}_{n=1}^{\infty}$ contains a convergent subsequence. But $\{Av_n\}_{n=1}^{\infty}$ is a Cauchy sequence, and a Cauchy sequence having a convergent subsequence is convergent. To see that $\{Av_n\}_{n=1}^{\infty}$ is Cauchy, write

$$Av_n = \sum_{k=1}^{n} \lambda_k \frac{\langle x, e_k \rangle}{\lambda_k - \lambda} e_k,$$

and observe that

$$\|Av_n - Av_m\|^2 = \sum_{k=n+1}^{m} \left| \lambda_k \frac{\langle x, e_k \rangle}{\lambda_k - \lambda} \right|^2 \leq \sup_{j \in \mathbb{N}} \left| \frac{\lambda_j}{\lambda_j - \lambda} \right|^2 \sum_{k=n+1}^{m} | \langle x, e_k \rangle |^2 \to 0.$$

This completes the proof. □

If an operator A is not necessarily compact but has a known spectral resolution in the form of an orthonormal series, then certain conclusions can be drawn, as shown in the following three theorems.

Theorem 2.4.9. *Let A be an operator on an inner product space having the form $Ax = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n$, where $\{e_n\}$ is an orthonormal sequence and $\{\lambda_n\}$ is a bounded sequence of nonzero complex numbers. Let $M := \operatorname{span}\{e_n : n \in \mathbb{N}\}$. Then $M^{\perp} = \ker(A)$.*

Proof. Recall that the following properties are equivalent of a vector $x \in X$:

(1) $x \in \ker(A)$;
(2) $\|Ax\|^2 = 0$;
(3) $\sum_{n=1}^{\infty} |\lambda_n \langle x, e_n \rangle|^2 = 0$;
(4) $\langle x, e_n \rangle = 0$ for all $n \in \mathbb{N}$.

□

Theorem 2.4.10. *Let A be an operator on an inner product space having the form $Ax = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n$, where $\{e_n\}$ is an orthonormal sequence and $\{\lambda_n\}$ is a bounded sequence of nonzero complex numbers. The orthonormal set $\{e_n\}_{n=1}^{\infty}$ is maximal if and only if* $\ker(A) = 0$.

Proof. By the preceding Theorem (2.4.9), $\ker(A) = \{0\}$ if and only if $M^{\perp} = \{0\}$. The condition $M^{\perp} = \{0\}$ is equivalent to the maximality of $\{e_n\}$, by the Orthonormal Basis Theorem (2.2.16). □

Theorem 2.4.11. *Let A be an operator on a Hilbert space such that $Ax = \sum_{n=1}^{\infty} \lambda_n \langle x, e_n \rangle e_n$, where $\{e_n\}$ is an orthonormal sequence $\{\lambda_n\}$ is a bounded sequence of nonzero complex numbers. If v is in the range of A, then one solution of the equation $Ax = v$ is $x := \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle v, e_n \rangle e_n$.*

Proof. Since v is in the range of A, $v = Az$ for some $z \in X$. Thus

$$\langle v, e_m \rangle = \langle Az, e_m \rangle = \left\langle \sum_{n=1}^{\infty} \lambda_n \langle z, e_n \rangle e_n, e_m \right\rangle = \lambda_m \langle z, e_m \rangle .$$

From this we have

$$\sum_{n=1}^{\infty} \left| \frac{1}{\lambda_n} \langle v, e_n \rangle \right|^2 = \sum_{n=1}^{\infty} | \langle z, e_n \rangle |^2 \leq \|z\|^2.$$

This implies the convergence of the series $x = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle v, e_n \rangle e_n$, by the Pythagorean Law. It follows that

$$Ax = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle v, e_n \rangle Ae_n = \sum_{n=1}^{\infty} \langle v, e_n \rangle e_n = \sum_{n=1}^{\infty} \lambda_n \langle z, e_n \rangle e_n = Az = v.$$

This completes the proof. □

Example 2.4.12. *Consider the operator A defined on $L^2[0,1]$ by the equation*

$$(Ax)(t) := \int_0^1 G(s,t)x(s)\ ds,$$

where

$$G(s,t) := \begin{cases} (1-s)t, & 0 \leq t \leq s \leq 1, \\ (1-t)s, & 0 \leq s \leq t \leq 1. \end{cases}$$

We first find all λ such that $Ax = \lambda x$. Observe that

$$(Ax)' = \frac{d}{dt} \left\{ \int_0^1 G(s,t)x(s)\ ds \right\}$$

$$= (1-t)tx(t) + \int_0^t \frac{\partial}{\partial t}(1-t)sx(s)\ ds - (1-t)tx(t) + \int_t^1 \frac{\partial}{\partial t}(1-s)tx(s)\ ds$$
$$= \int_0^t -sx(s)\ ds + \int_t^1 (1-s)x(s)\ ds$$

and

$$\begin{aligned}(Ax)'' &= \frac{d}{dt}\left\{\int_0^t -sx(s)\ ds + \int_t^1 (1-s)x(s)\ ds\right\} \\ &= -tx(t) + \int_0^t 0\ ds - (1-t)x(t) + \int_0^t 0\ ds \\ &= x(t).\end{aligned}$$

Hence finding the eigenvalues and eigenfunctions of A is equivalent to solving the BVP

$$\begin{cases} x'' - \lambda x = 0, \\ x'(0) + x(0) = 0, \\ x'(1) - x(1) = 0. \end{cases}$$

For $\lambda < 0$, the general solution to the BVP is

$$x(t) := c_1 \sin\sqrt{\lambda}t + c_2 \cos\sqrt{\lambda}t.$$

We find

$$x'(t) = c_1\lambda\cos\sqrt{\lambda}t - c_2\lambda\sin\sqrt{\lambda}t.$$

Applying the BC's, we have

$$x'(0) - x(0) = c_1\lambda + c_2 = 0$$

and

$$x'(1) - x(1) = (c_1\lambda - c_2)\cos\sqrt{\lambda} - (c_2\lambda + c_1)\sin\sqrt{\lambda} = 0.$$

For compact operators that are not self–adjoint, there is still a useful canonical form that can be exploited.

Theorem 2.4.13 (Singular Value Decomposition for Compact Operators)**.** *Every compact operator A on a separable Hilbert space is expressible in the form*

$$Ax = \sum_{n=1}^{\infty} \langle x, u_n \rangle v_n,$$

in which $\{u_n\}$ is an orthonormal basis for the space and $\{v_n\}$ is an orthogonal sequence converging to zero.

Proof. Note that the operator A^*A is compact and Hermitian since, for any $x, y \in X$, we see that

$$\langle A^*Ax, y\rangle = \langle x, (A^*A)^*y\rangle = \langle x, A^*A^{**}y\rangle = \langle x, A^*Ay\rangle .$$

Moreover, its eigenvalues are nonnegative, for if $A^*Ax = \mu x$, then

$$0 \le \|Ax\|^2 \langle Ax, Ax\rangle = \langle x, A^*Ax\rangle = \langle x, \mu x\rangle = \mu x, x.$$

Applying the spectral theorem to A^*A, we obtain

$$A^*Ax = \sum_{n=1}^{\infty} \lambda_n^2 \langle x, u_n\rangle u_n$$

for some orthonormal basis $\{u_n\}$ and sequence $\{\lambda_n^2\}$ such that $\lim_{n\to\infty} \lambda_n^2 = 0$. Note that in this representation, each nonzero eigenvalue λ_n^2 is repeated a number of times equal to its geometric multiplicity. Define $v_n := Au_n$. Then we have

$$\langle v_m, v_n\rangle = \langle Au_m, Au_n\rangle = \langle u_m, A^*Au_n\rangle = \left\langle u_m, \lambda_n^2 u_n\right\rangle = \lambda_n^2 \delta_{mn}.$$

Thus $\{v_n\}$ is orthogonal, and $\|v_n\| = \lambda_n \to 0$. Since $\{u_n\}$ is a basis, we have, for arbitrary $x \in X$,

$$x = \sum_{n=1}^{\infty} \langle x, u_n\rangle\, u_n.$$

Consequently,

$$Ax = \sum_{n=1}^{\infty} \langle x, u_n\rangle\, Au_n = \sum_{n=1}^{\infty} \langle x, u_n\rangle\, v_n.$$

□

Definition 2.4.14 (Hilbert–Schmidt Operator)**.** *A* **Hilbert–Schmidt operator** *is a compact operator A on a Hilbert space such that*

$$\sum_{\alpha} \|Au_\alpha\|^2 < \infty,$$

for some orthonormal basis $\{u_\alpha\}$.

Theorem 2.4.15. *Let $\{u_\alpha\}$ and $\{v_\beta\}$ be two orthonormal bases for a Hilbert space X. Every linear operator A on the space satisfies*

$$\sum_{\alpha} \|Au_\alpha\|^2 = \sum_{\beta} \|Av_\beta\|^2.$$

Proof. By the Orthonormal Basis Theorem (2.2.16) and the Parseval Identity, we have

$$\begin{aligned}\sum_{\alpha} \|Au_\alpha\|^2 &= \sum_{\alpha}\sum_{\beta} |\langle Au_\alpha, v_\beta\rangle|^2 = \sum_{\beta}\sum_{\alpha} |\langle Au_\alpha, v_\beta\rangle|^2 \\ &= \sum_{\beta}\sum_{\alpha} |\langle u_\alpha, A^*v_\beta\rangle|^2 = \sum_{\beta} \|A^*v_\beta\|^2.\end{aligned}$$

Since $\{u_\alpha\}$ and $\{v_\beta\}$ may switch roles in this calculation, we obtain $\sum_\beta \|Av_\beta\|^2 = \sum_\beta \|A^*v_\beta\|^2$. By combining these equations, the conclusion follows. □

2.5. **Sturm–Liouville Theory.** In this section we examine differential equations using Hilbert space theory. Note that differential operators and integral operators are inverses to each other. We find that a differential operator is usually ill–behaved, whereas the corresponding integral operator may be well–behaved, even to the point of being compact. Thus we often try to recast a differential equation as an equivalent integral equation in the hope that the transformed problem will be less troublesome.

Definition 2.5.1 (Sturm–Liouville Operator)**.** *The* **Sturm–Liouville operator** *A is defined by*

$$(Ax)(t) := [p(t)x'(t)]' + q(t)x(t), \qquad \text{i.e.,} \qquad Ax := (px')' + qx,$$

where $x \in \mathcal{C}^2([a,b],\mathbb{C})$, $p \in \mathcal{C}^1([a,b],\mathbb{R})$, and $q \in \mathcal{C}([a,b],\mathbb{R})$. Let $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$, $i,j = 1,2$ be such that

$$p(a)(\beta_{11}\beta_{22} - \beta_{12}\beta_{21}) = p(b)(\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}).$$

Let X be the subspace of $L^2[a,b]$ consisting of all twice continuously differentiable functions x such that

$$\alpha_{11}x(a) + \alpha_{12}x'(a) + \beta_{11}x(b) + \beta_{12}x'(b) = 0,$$
$$\alpha_{21}x(a) + \alpha_{22}x'(a) + \beta_{21}x(b) + \beta_{22}x'(b) = 0.$$

Further assume that $\beta_{11}\beta_{22} \neq \beta_{12}\beta_{21}$ or $\alpha_{11}\alpha_{22} \neq \alpha_{12}\alpha_{21}$.

Theorem 2.5.2. *The Sturm–Liouville operator A is a Hermitian operator on X.*

Proof. Let $x, y \in X$. Recall that we want to show that $\langle Ax, y\rangle = \langle x, Ay\rangle$. We find

$$\begin{aligned}
\langle Ax, y\rangle - \langle x, Ay\rangle &= \int_a^b [\overline{y}Ax - xA\overline{y}] = \int_a^b [\overline{y}(px')' + \overline{y}qx - x(p\overline{y}')' - xq\overline{y}] \\
&= \int_a^b [\overline{y}(px')' - x(p\overline{y}')'] \\
&= \int_a^b [\overline{y}(px')' + \overline{y}'px' - x(p\overline{y}')' - x'p\overline{y}'] \\
&= \int_a^b [px'\overline{y} - px\overline{y}']' = [px'\overline{y} - px\overline{y}']\Big|_a^b \\
&= p(b)[x'(b)\overline{y}(b) - x(b)\overline{y}'(b)] - p(a)[x'(a)\overline{y}(a) - x(a)\overline{y}'(a)]
\end{aligned}$$

Define

$$W(t) := \begin{bmatrix} x(t) & \overline{y}(t) \\ x'(t) & \overline{y}'(t) \end{bmatrix}.$$

Then note that

$$\langle Ax, y\rangle - \langle x, Ay\rangle = -p(b)\det W(b) + p(a)\det W(a).$$

Put also

$$\alpha := \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}, \qquad \beta := \begin{bmatrix} \beta_{11} & \beta_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix}.$$

Note that p is such that $p(a)\det\beta = p(b)\det\alpha$. The fact that $x, y \in X$ gives $\alpha W(a) + \beta W(b) = 0$. This yields $\det\alpha\det W(a) = \det(-\beta)\det W(b)$. Note $\det(-\beta) = \det\beta$ since β is of even order. Multiplying this by $p(b)$ gives

$$p(b)\det\alpha\det W(a) = p(b)\det\beta\det W(b).$$

By the previous analysis, this is

$$p(a)\det\beta\det W(a) = p(b)\det\beta\det W(b).$$

If $\det\beta \neq 0$, we have

$$p(a)\det W(a) = p(b)\det W(b).$$

If $\det\alpha \neq 0$, we reach a similar conclusion. ∎

Lemma 2.5.3. *A second–order linear differential equation*

$$a(t)x''(t) + b(t)x'(t) + c(t)x(t) = d(t), \quad a \le t \le b$$

can be put into the form of a Sturm–Liouville equation $(px')' + qx = f$, provided that the functions a, b, c are continuous and $a(t) \neq 0$ for all $t \in [a,b]$.

Proof. Multiplying the differential equation by the integrating factor $\frac{1}{a}e^{\int b/a\ dt}$ gives

$$x''e^{\int b/a} + \frac{b}{a}x'e^{\int b/a} + \frac{c}{a}xe^{\int b/a} = \frac{d}{a}e^{\int b/a},$$

or

$$\left(x'e^{\int b/a}\right)' + \frac{c}{a}e^{\int b/a}x = \frac{d}{a}e^{\int b/a}.$$

Letting

$$p := e^{\int b/a}, \qquad q := \frac{c}{a}e^{\int b/a}, \qquad f := \frac{d}{a}e^{\int b/a}$$

completes the proof. □

Example 2.5.4. *Consider*

$$\begin{cases} Ax = -x'', \\ x(0) = 0, \quad x(\pi) = 0. \end{cases}$$

Note that in this case $p \equiv -1$ and $q \equiv 0$. Solving for the general solution to the BVP gives

$$x(t) = c_1 \sin\sqrt{\lambda}t + c_2 \cos\sqrt{\lambda}t.$$

Applying the BC's, we find $c_2 \equiv 0$ and $c_1 \sin\sqrt{\lambda}\pi = 0$, where the latter equation holds if and only if $\sqrt{\lambda} \in \mathbb{N}$. Hence the eigenvalues of the operator A are $\lambda_n = n^2$, $n \in \mathbb{N}$, and the associated eigenfunctions are $x_n = \sin n^2 t$.

The proceeding theorem illustrates one case of the Sturm–Liouville problem. We take $p \equiv 1$ in the differential operator and let $\beta_{11} = \beta_{12} = \alpha_{21} = \alpha_{22} = 0$. We assume that $|\alpha_{11}| + |\alpha_{12}| > 0$ and $|\beta_{21}| + |\beta_{22}| > 0$.

Our goal is to develop a method for solving the equation $Ax = y$, where y is a given function, and x is to be determined. We find a right inverse of A, that is, an operator B such that $AB = I$, and give $x = By$ as a solution to the problem. It will turn out that the spectral theorem is applicable to B.

Assume that there exist functions u, v such that

$$\begin{aligned} u'' &= qu, \qquad \beta_{21}u(b) + \beta_{22}u'(b) = 0, \\ v'' &= qv, \qquad \alpha_{11}v(a) + \alpha_{12}v'(a) = 0, \\ u'(a)&v(a) - u(a)v'(a) = 1. \end{aligned}$$

We see that $u \neq 0$ and $v \neq 0$, for the LHS of the third equation is the Wronskian of u and v evaluated at a.

We find u and v be solving two initial value problems. We proceed as follows. Find u_0 and v_0 such that

$$\begin{aligned} u_0'' &= qu_0, \qquad u_0(b) = 0, \qquad u_0'(b) = 1, \\ v_0'' &= qv_0, \qquad v_0(a) = 1, \qquad v_0'(a) = 0. \end{aligned}$$

The u and v required will then be suitable linear combinations of u_0 and v_0.

Next, we see that for all s,

$$u'(s)v(s) - u(s)v'(s) = 1.$$

This is true because the LHS is prescribed to take the value 1 at $s = a$ and is constant. Indeed,

$$\frac{d}{ds}[u'v - uv'] = u''v + u'v' - uv'' - u'v' = quv - uqv = 0.$$

We now construct a **Green's function** g for the problem. We define

$$g(s,t) := \begin{cases} u(s)v(t), & a \le t \le s \le b, \\ u(t)v(s), & a \le s \le t \le b. \end{cases}$$

Recall that the operator in this case ($p \equiv 1$) is defined by

$$Ax = x'' - qx$$

and the domain of A is the $L^2[a,b]$ closure of the set of all twice continuously differentiable functions x such that

$$\alpha_{11}x(a) + \alpha_{12}x'(a) = \beta_{21}x(b) + \beta_{22}x'(b) = 0.$$

Theorem 2.5.5. *A right inverse of the differential operator $Ax := x'' - qx$ is the operator B defined by*

$$(By)(s) := \int_a^b g(s,t)y(t)\ dt.$$

Proof. We show that $AB \equiv I$. Let $y \in \mathcal{C}[a,b]$ and put $x := By$. We show first that $Ax = y$. From the equation

$$\begin{aligned} x(s) &= \int_a^b g(s,t)y(t)\ dt \\ &= \int_a^s u(s)v(t)y(t)\ dt + \int_s^b u(t)v(s)y(t)\ dt \\ &= u(s)\int_a^s v(t)y(t)\ dt + v(s)\int_s^b u(t)y(t)\ dt, \end{aligned}$$

we have

$$\begin{aligned} x'(s) &= u'(s)\int_a^s v(t)y(t)\ dt + u(s)v(s)y(s) + v'(s)\int_s^b u(t)y(t)\ dt - v(s)u(s)y(s) \\ &= u'(s)\int_a^s v(t)y(t)\ dt + v'(s)\int_s^b u(t)y(t)\ dt. \end{aligned}$$

Differentiating again, we obtain

$$\begin{aligned} x''(s) &= u''(s)\int_a^s v(t)y(t)\ dt + u'(s)v(s)y(s) + v''(s)\int_s^b u(t)y(t)\ dt - v'(s)u(s)y(s) \\ &= q(s)u(s)\int_a^s v(t)y(t)\ dt + q(s)v(s)\int_s^b u(t)y(t)\ dt + y(s)[u'(s)v(s) - u(s)v'(s)] \\ &= q(s)x(s) + y(s), \end{aligned}$$

by definition of x and since $u'(s)v(s) - u(s)v'(s) \equiv 1$. Thus $x'' - qx \equiv y$, or $Ax = y$, as asserted.

We now show that $x \in X$, that is, that x satisfies the boundary conditions. We have, from above,

$$x(a) = v(a)\int_a^b u(t)y(t)\ dt = cv(a),$$

and

$$x'(a) = v'(a) \int_a^b u(t)y(t)\ dt = cv'(a),$$

where

$$c := \int_a^b u(t)y(t)\ dt.$$

Hence,

$$\alpha_{11}x(a) + \alpha_{12}x'(a) = \alpha_{11}cv(a) + \alpha_{12}cv'(a) = 0.$$

Similarly we find that

$$\beta_{21}x(b) + \beta_{22}x'(b) = 0.$$

This completes the proof. □

Corollary 2.5.6. *If the homogeneous boundary value problem has only the trivial solution, then B is also a left inverse of A.*

Proof. Let $x \in X$, $y = Ax$, and $By = z$. By the previous Theorem (2.5.5), we have $y = ABy = Az$, so that $z \in X$. Thus $x - z \in X$ and $A(x - z) = 0$. Further, $x - z = 0$, so that $x = By = BAx$. □

Corollary 2.5.7. *The operator B is Hermitian.*

Now we may apply the Spectral Theorem to the operator B. Note that B is compact by Theorem (2.3.17). By the Spectral Theorem, there exists an orthonormal sequence $\{u_n\}$ in $L^2[a, b]$ and a sequence of real numbers $\{\lambda_n\}$ such that

$$By = \sum_{n=1}^{\infty} \lambda_n \langle y, u_n \rangle u_n.$$

Since $Bu_k = \lambda_k u_k$, we have $u_k = \lambda_k A u_k$, and u_k satisfies the boundary conditions. This equation shows that u_k is an eigenvector of A corresponding to the eigenvalue $1/\lambda_k$. Since $\lim_{k\to\infty} \lambda_k = 0$, $\lim_{k\to\infty} 1/\lambda_k = \infty$. Consequently, a solution to the problem $Ax = y$, where y is given and x must satisfy the boundary conditions, is

$$x = By = \sum_{n=1}^{\infty} \lambda_n \langle y, u_n \rangle u_n.$$

Example 2.5.8. *Consider the BVP*

$$\begin{cases} Ax = x'' + x = y, \\ x'(0) = x(\pi) = 0. \end{cases}$$

We solve this problem by means of a Green's function. Note here that $p \equiv 1$, $q \equiv -1$, and we solve the following two IVPs:

$$\begin{cases} u_0'' + u_0 = 0, \\ u_0(\pi) = 1, \quad u_0'(\pi) = 0, \end{cases} \qquad \begin{cases} v_0'' + v_0 = 0, \\ v_0(0) = 0, \quad v_0'(0) = 1. \end{cases}$$

Solving for u_0 and v_0, we obtain

$$u_0(t) = \sin t, \qquad v_0(t) = \cos t.$$

Note that taking

$$u(t) := \sin t, \qquad v(t) := \cos t$$

satisfies the boundary conditions. Thus the Green's function g for this BVP is

$$g(s,t) = \begin{cases} \sin s \cos t, & 0 \le t \le s \le \pi, \\ \sin t \cos s, & 0 \le s \le t \le \pi. \end{cases}$$

Hence, by Theorem (2.5.5)*, a solution x to the BVP is given by*

$$x(t) := \int_0^\pi g(s,t)y(t)\ dt = \sin s \int_0^\pi \cos t y(t)\ dt + \cos s \int_s^\pi \sin t y(t)\ dt.$$

Our next task is to find a method to determine a Green's function for the more general Sturm–Liouville problem. Recall that the differential equation and its boundary conditions are as follows:

$$\begin{cases} Ax := (px')' + qx = y, \quad x \in \mathcal{C}^2[a,b], \\ \alpha_{11}x(a) + \alpha_{12}x'(a) + \beta_{11}x(b) + \beta_{12}x'(b), \\ \alpha_{21}x(a) + \alpha_{22}x'(a) + \beta_{21}x(b) + \beta_{22}x'(b). \end{cases}$$

We determine a function g defined on $[a,b] \times [a,b]$.

Theorem 2.5.9. *The Green's function for the general Sturm–Liouville problem is characterized by the following five properties:*

(1) g is continuous throughout $[a,b] \times [a,b]$;
(2) $\frac{\partial g}{\partial s}$ is continuous throughout $a < s < t < b$ and $a < t < s < b$;
(3) For all $t \in [a,b]$, $g(\cdot,t)$ satisfies the boundary conditions;
(4) $Ag(\cdot,t) = 0$ throughout $a < s < t < b$ and $a < t < s < b$;
(5) $\lim_{t\to s+} \frac{\partial g}{\partial s}(s,t) - \lim_{t \to s-} \frac{\partial g}{\partial s}(s,t) = -\frac{1}{p(s)}$.

Proof. We again take $y \in \mathcal{C}[a,b]$ and define

$$x(s) := \int_a^b g(s,t)y(t)\ dt.$$

We show that $x \in \mathrm{dom}A$ and $Ax = y$. Note that the domain of A is the set of all twice continuously differentiable functions that satisfy the boundary conditions. Since

$$x(s) = \int_a^s g(s,t)y(t)\ dt + \int_s^b g(s,t)y(t)\ dt,$$

we have

$$\begin{aligned} x'(s) &= \int_a^s g_s(s,t)y(t)\ dt + g(s,s)y(s) + \int_s^b g_s(s,t)y(t)\ dt - g(s,s)y(s) \\ &= \int_a^s g_s(s,t)y(t)\ dt + \int_s^b g_s(s,t)y(t)\ dt. \end{aligned}$$

Thus it follows

$$\begin{aligned} x(a) &= \int_a^b g(a,t)y(t)\ dt, \qquad & x(b) &= \int_a^b g(b,t)y(t)\ dt, \\ x'(a) &= \int_a^b g_s(a,t)y(t)\ dt, \qquad & x'(b) &= \int_a^b g_s(b,t)y(t)\ dt. \end{aligned}$$

Note that any linear combination of $x(a), x(b), x'(a), x'(b)$ is obtained by an integration of the corresponding linear combination of $g(a,t), g(b,t), g_s(a,t), g_s(b,t)$. Since $g(\cdot,t)$ satisfies the boundary conditions, so does x. Differentiating x again, we obtain

$$\begin{aligned} x''(s) &= \int_a^s g_{ss}(s,t)y(t)\ dt + g_s(s,s-)y(s) + \int_s^b g_{ss}(s,t)y(t)\ dt - g_s(s,s+)y(s) \\ &= \frac{y(s)}{p(s)} + \int_a^b g_{ss}(s,t)y(t)\ dt. \end{aligned}$$

We now verify that $Ax = y$. Recall that

$$Ax = (px')' + qx = p'x' + px'' + qx.$$

Thus

$$\begin{aligned} (Ax)(s) &= p'(s)\int_a^b g_s(s,t)y(t)\ dt + y(s) + p(s)\int_a^b g_{ss}(s,t)y(t)\ dt + q(s)\int_a^b g(s,t)y(t)\ dt \\ &= y(s) + \int_a^b \frac{\partial}{\partial s}[p(s)g_s(s,t) + q(s)g(s,t)]y(t)\ dt \\ &= y(s) + \int_a^b Ag(s,t)y(t)\ dt \\ &= y(s), \end{aligned}$$

since $Ag(\cdot,t) = 0$ almost everywhere. This completes the proof. □

Example 2.5.10. *We find the Green's function for this Sturm–Liouville problem:*

$$\begin{cases} Ax = x'' = y, & x \in \mathcal{C}^2[0,1], \\ x(0) = 0, & x'(0) = 0. \end{cases}$$

Note that the general solution to this problem is

$$x(s) = c_1 + c_2 s.$$

Since g solves the differential equation, we take

$$g(s,t) := c_1(t) + c_2(t)s.$$

First consider $0 < s < t < 1$. Since $g(t,\cdot)$ must satisfy the boundary conditions, we have

$$g(0,t) = c_1(t) = 0, \quad \frac{\partial g}{\partial s}(0,t) = c_2(t) = 0.$$

Consider next $0 < t < s < 1$. Note again that $g(t,\cdot)$ must be linear. Thus

$$g(s,t) := \alpha(t) + \beta(t)s.$$

Continuity of g on the diagonal implies that

$$\alpha(t) + \beta(t)t = 0,$$

and thus

$$g(s,t) = -\beta(t)t + \beta(t)s = \beta(t)(s-t).$$

The condition

$$g_s(s,s+) - g_s(s,s-) = -\frac{1}{p(s)}$$

implies

$$0 - \beta(t) = -1.$$

Hence $\beta(t) = 1$. Thus the Green's function $g(s,t)$ is given by

$$g(s,t) = \begin{cases} 0, & 0 \le s \le t \le 1, \\ s-t, & 0 \le t \le s \le 1, \end{cases}$$

and the solution to the BVP is given by

$$x(s) = \int_0^s (s-t)y(t)\ dt.$$

Example 2.5.11. *We find the Green's function for the problem*

$$\begin{cases} Ax = x'' - x' - 2x = y, & x \in \mathcal{C}^2[0,1], \\ x(0) = 0, \quad x(1) = 0. \end{cases}$$

We set

$$g(s,t) = \begin{cases} u(s)v(t), & 0 \le s \le t \le 1, \\ u(t)v(s), & 0 \le t \le s \le 1, \end{cases}$$

and try to determine the functions u and v. The general solution to the homogeneous ODE is given by

$$x(s) := c_1 e^{-s} + c_2 e^{2s}.$$

The solution satisfying the condition $x(0) = 0$ is

$$u(s) = c_1(e^{-s} - e^{2s}),$$

and the solution satisfying the condition $x(1) = 0$ is

$$v(s) = c_2(e^{2s} - e^3 e^{-s}).$$

It may be shown that $g(s,t)$ satisfies the first four requirements of Theorem (2.5.9). We determine c_1 and c_2. Note

$$g_s(s,s+) - g_s(s,s-) = u'(s)v(s) - u(s)v'(s) = 3c_1c_2(1-e^3)e^s.$$

In this problem $p(s) = e^{-s}$, because

$$p(s) := e^{\int -1/1\ ds} = e^{-s}.$$

Thus condition (5) in Theorem (2.5.9) implies that we choose c_1, c_2 such that

$$c_1c_2 = -\frac{1}{3(1-e^3)}.$$

Thus

$$g(s,t) = \begin{cases} -\frac{1}{3(1-e^3)}(e^{-s} - e^{2s})(e^{2t} - e^{3-t}), & 0 \le s \le t \le 1, \\ -\frac{1}{3(1-e^3)}(e^{-t} - e^{2t})(e^{2s} - e^{3-s}), & 0 \le t \le s \le 1. \end{cases}$$

Example 2.5.12. *We find the Green's function for this Sturm–Liouville problem:*

$$\begin{cases} Ax = x'' + 9x = y, & x \in \mathcal{C}^2[0,\pi/2], \\ x(0) = 0, \quad x(\pi/2) = 0. \end{cases}$$

We again write

$$g(s,t) = \begin{cases} u(s)v(t), & 0 \le s \le t \le 1, \\ u(t)v(s), & 0 \le t \le s \le 1. \end{cases}$$

The general solution is

$$x(s) = c_1 \sin 3s + c_2 \cos 3s.$$

The solution satisfying $x(0) = 0$ *is*

$$u(s) = c_1 \sin 3s,$$

and the solution satisfying $x(\pi/2) = 0$ *is*

$$v(s) = c_2 \cos 3s.$$

We have

$$g_s(s, s+) - g_s(s, s-) = u'(s)v(s) - u(s)v'(s) = 3c_1c_2.$$

Here $p(s) = 1$, *so we must take* $c_1c_2 = -1/3$. *Hence the Green's function is*

$$g(s,t) = \begin{cases} -\frac{1}{3} \sin 3s \cos 3t, & 0 \le s \le t \le 1, \\ -\frac{1}{3} \sin 3t \cos 3s, & 0 \le t \le s \le 1. \end{cases}$$

3. Calculus in Banach Spaces

3.1. **The Fréchet Derivative.**

Definition 3.1.1 (Fréchet Derivative)**.** *Let X and Y be normed linear spaces and let $D \subseteq X$ be an open set. Let $f : D \to Y$ be a mapping and let $x \in D$. If there is a bounded linear map $A : X \to Y$ such that*

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Ah\|_Y}{\|h\|_X} = 0,$$

then f is said to be **Fréchet differentiable** *at x. Moreover, A is called the* **Fréchet derivative** *of f at x.*

Theorem 3.1.2. *If f is differentiable at x with Fréchet derivative A, then A is uniquely defined.*

Proof. Suppose that A_1 and A_2 are both linear maps satisfying

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - A_i h\|}{\|h\|} = 0,$$

for $i = 1, 2$. Fix $\epsilon > 0$. By definition, there exists $\delta > 0$ such that whenever $\|h\| < \delta$, we have

$$\|f(x+h) - f(x) - A_i h\| < \epsilon\|h\|,$$

$i = 1, 2$. By the triangle inequality,

$$\|A_1 h - A_2 h\| = \|(f(x+h) - f(x) - A_2 h) - (f(x+h) - f(x) - A_1 h)\| \leq 2\epsilon\|h\|$$

for all $\|h\| < \delta$. By homogeneity, this inequality is true for all h. Thus $\|A_1 - A_2\| \leq 2\epsilon$. Since ϵ was arbitrary, it follows $A_1 = A_2$. □

If f is differentiable at x, its derivative, denoted by A in the preceding definition, is usually denoted by $f'(x)$. Note that with this notation, $f'(x) \in \mathcal{L}(X, Y)$. This is *not* saying $f' \in \mathcal{L}(X, Y)$. Rather, $f' \in \mathcal{L}(X, \mathcal{L}(X, Y))$.

Theorem 3.1.3. *If f is bounded in a neighborhood of x and if a linear map A has the property*

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - Ah\|}{\|h\|} = 0,$$

then A is a bounded linear map, that is, A is the Fréchet derivative of f at x.

Proof. By the assumption, there exists $\delta > 0$ such that whenever $\|h\| \leq \delta$, we have

$$\|f(x+h)\| \leq M < \infty$$

and

$$\|f(x+h) - f(x) - Ah\| \leq \|h\|.$$

Then, for all $\|h\| \leq \delta$, we have

$$\|Ah\| \leq \|f(x+h) - f(x)\| + \|h\| \leq 2M + \delta.$$

Hence, for all $\|u\| \leq 1$, $\|\delta u\| \leq \delta$, and we have

$$\|Au\| = \frac{1}{\delta}\|A(\delta u)\| \leq \frac{2M + \delta}{\delta} = \frac{2M}{\delta} + 1.$$

Thus $\|A\| \leq (2M + \delta)/\delta$, which completes the proof. $\square$

Example 3.1.4. *Let $X = Y = \mathbb{R}$. Let f be a function whose derivative in the usual sense at x is λ. Then the Fréchet derivative of f at x is the linear map $Ah = \lambda h$, because*

$$\lim_{h\to 0} \frac{|f(x+h) - f(x) - \lambda h|}{|h|} = \lim_{h\to 0} \left| \frac{f(x+h) - f(x)}{h} - \lambda \right| = 0.$$

Note here that the derivative at the **point** *x is the map $Ah = \lambda h$.*

Example 3.1.5. *Let X and Y be arbitrary normed linear spaces. Define $f : X \to Y$ be $f(x) = y_0$, where $y_0 \in Y$ is a fixed element. That is, f is a constant map. Then $f'(x) = 0$, where 0 denotes the zero mapping of Y, because*

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - 0\|}{\|h\|} = \lim_{h\to 0} \frac{\|0\|}{\|h\|} = 0.$$

Example 3.1.6. *Let X and Y be normed linear spaces and let $f \in \mathcal{L}(X, Y)$. Then, for any $x \in X$, $f'(x) = f$, for we see that*

$$\lim_{h\to 0} \frac{\|f(x+h) - f(x) - f(h)\|}{\|h\|} = \lim_{h\to 0} \frac{\|f(h) - f(h)\|}{\|h\|} = 0.$$

Observe that the equation $f' = f$ is **not** *true. Rather, $f(h) = f'(x)h$.*

Theorem 3.1.7. *If f is Fréchet differentiable at $x \in X$, then f is continuous at x.*

Proof. Let $A = f'(x)$. Note that $A \in \mathcal{L}(X, Y)$. Fix $\epsilon > 0$. Then there exists δ_1 such that $\delta_1 < \epsilon/(1 + \|A\|)$. On the other hand, there exists δ_2 such that for all $\|h\| < \delta_2$, we have

$$\|f(x+h) - f(x) - Ah\| < \|h\|.$$

Choose $\delta := \min\{\delta_1, \delta_2\}$. Then, by the triangle inequality, we see for all $\|h\| < \delta$ that

$$\begin{aligned} \|f(x+h) - f(x)\| &\leq \|f(x+h) - f(x) - Ah\| + \|Ah\| \\ &< \|h\| + \|Ah\| \leq \|h\| + \|A\|\|h\| \\ &< \delta(1 + \|A\|) < \epsilon, \end{aligned}$$

which completes the proof. $\square$

Theorem 3.1.8. *Let $f : \mathbb{R}^n \to \mathbb{R}$. If each of the partial derivatives $D_i f = \partial f/\partial x_i$ exists in a neighborhood of x and is continuous at x, then $f'(x)$ exists, and*

$$f'(x)h = \sum_{i=1}^{n} D_i f(x) \cdot h_i, \quad h = (h_1, h_2, \ldots, h_n)^\top \in \mathbb{R}^n.$$

Proof. We show that

$$\lim_{h\to 0} \frac{1}{\|h\|} \left\{ f(x+h) - f(x) - \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x) \right\} = 0.$$

Define the vectors $v^{(i)}, i = 0, 1, \ldots, n$ by

$$v^{(0)} := x, \quad v^{(i)} := v^{(i-1)} + h_i e^{(i)}, \ i = 1, 2, \ldots, n,$$

where $e^{(i)}$ denotes the $i-$th standard basis vector in $\mathbb{R}^n$. Note that the vectors $v^{(i)}$ and $v^{(i-1)}$ differ in only one coordinate, for $i = 1, 2, \dots, n$. Thus we have

$$f(x+h) - f(x) = f(v^{(n)}) - f(v^{(0)}) = \sum_{i=1}^{n} [f(v^{(i)}) - f(v^{(i-1)})].$$

By the mean value theorem for functions of one variable,

$$f(v^{(i)}) - f(v^{(i-1)}) = f(v^{(i-1)} + h_i e^{(i)}) - f(v^{(i-1)}) = h_i \frac{\partial f}{\partial x_i}(v^{(i-1)} + \theta_i h_i e^{(i)}),$$

where $0 < \theta_i < 1$, $i = 1, 2, \dots, n$. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\frac{1}{\|h\|}\left| f(x+h) - f(x) - \sum_{i=1}^{n} h_i \frac{\partial f}{\partial x_i}(x)\right| &= \frac{1}{\|h\|}\left|\sum_{i=1}^{n} h_i \left[\frac{\partial f}{\partial x_i}(v^{(i-1)} + \theta_i h_i e^{(i)}) - \frac{\partial f}{\partial x_i}(x)\right]\right| \\ &\leq \frac{1}{\|h\|}\|h\| \sqrt{\sum_{i=1}^{n}\left[\frac{\partial f}{\partial x_i}(v^{(i-1)} + \theta_i h_i e^{(i)}) - \frac{\partial f}{\partial x_i}(x)\right]^2},\end{aligned}$$

which tends to zero as $\|h\| \to 0$ by the continuity of $\frac{\partial f}{\partial x_i}$ at x. We see that

$$\|v^{(i-1)} + \theta_i h_i e^{(i)} - x\| = \|(h_1, h_2, \dots, h_{i-1}, \theta_i h_i, 0, 0, \dots, 0)^\top\| \leq \|h\|. \qquad \square$$

Theorem 3.1.9. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and let $f_1, \dots, f_m$ be the component functions of f. If all partial derivatives $\partial f_i / \partial x_j$, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$, exist in a neighborhood of x and are continuous at x, then $f'(x)$ exists, and*

$$(f'(x)h)_i = \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(x) \cdot h_j, \quad h = (h_1, h_2, \dots, h_n)^\top \in \mathbb{R}^n.$$

That is, the Fréchet derivative of f is given by the Jacobian matrix J of f at x, $J_{ij} = \frac{\partial f_i}{\partial x_j}(x)$.

Proof. By definition of the Euclidean norm,

$$\frac{1}{\|h\|^2}\|f(x+h) - f(x) - Jh\|^2 = \frac{1}{\|h\|^2}\sum_{i=1}^{m}\left| f_i(x+h) - f_i(x) - \sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}(x) \cdot h_j\right|^2.$$

Note that each of the m terms on the RHS converges to zero as $h \to 0$, by the preceding Theorem (3.1.8). This completes the proof. $\square$

Example 3.1.10. *Let L be a bounded linear operator on a real Hilbert space X. Define $F : X \to \mathbb{R}$ by the equation $F(x) := \langle x, Lx\rangle$. In order to see whether F is differentiable at x, write*

$$\begin{aligned}F(x+h) - F(x) &= \langle x+h, Lx + Lh\rangle - \langle x, Lx\rangle \\ &= \langle x, Lh\rangle + \langle h, Lx\rangle + \langle h, Lh\rangle.\end{aligned}$$

Since the derivative is a linear map, we guess that A should be $Ah = \langle x, Lh\rangle + \langle h, Lx\rangle$. With this choice,

$$|Ah| \leq 2\|x\|\|L\|\|h\|,$$

so that

$$\|A\| \leq 2\|x\|\|L\|.$$

Thus A is a bounded linear functional. Moreover,

$$|F(x+h) - F(x) - Ah| = |\langle h, Lh\rangle| \leq \|L\|\|h\|^2 = o(h).$$

This shows that $A = F'(x)$. Note that

$$Ah = \langle L^*x + Lx, h\rangle.$$

3.2. **The Chain Rule and Mean Value Theorems.** We continue to work with a function $f : D \to Y$, where $D \subset X$ is an open set in the normed linear space X, and Y is another normed linear space. In the next theorem, we have another mapping g defined on an open set in Y and taking values in a third normed linear space, say Z.

Theorem 3.2.1 (The Chain Rule)**.** *If f is differentiable at x and g is differentiable at $f(x)$, then $g \circ f$ is differentiable at x, and*

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x).$$

Proof. Define $F := g \circ f$, $A := f'(x)$, $y := f(x)$, $B := g'(y)$, and

$$\begin{aligned} o_1(h) &:= f(x+h) - f(x) - Ah, \\ o_2(k) &:= g(y+k) - g(k) - Ak, \\ \phi(h) &:= Ah + o_1(h). \end{aligned}$$

We show that $F'(x) = BA$. We calculate

$$\begin{aligned} F(x+h) - F(x) - BAh &= g(f(x+h)) - g(f(x)) - BAh \\ &= g(f(x) + Ah + o_1(h)) - g(y) - BAh \\ &= g(y + \phi(h)) - g(y) - BAh \\ &= g(y) + B\phi(h) + o_2(\phi(h)) - g(y) - BAh \\ &= B[Ah + o_1(h)] + o_2(\phi(h)) - BAh \\ &= B(o_1(h)) + o_2(\phi(h)). \end{aligned}$$

Thus it follows

$$\|F(x+h) - F(x) - BAh\| = \|B(o_1(h)) + o_2(\phi(h))\| \leq \|B\|\|o_1(h)\| + \|o_2(\phi(h))\|.$$

Since the sum of two $o(h)$ functions is $o(h)$, the conclusion follows. $\square$

The mean value theorem of calculus does not have an exact analogue for mappings between general normed linear spaces. Even for functions $f : \mathbb{R} \to X$, the expected mean value theorem fails.

Example 3.2.2. *Define $f : \mathbb{R} \to \mathbb{R}^2$ by $f(t) := (\cos t, \sin t)$. Note that there exists no $t \in (0, 2\pi)$ such that*

$$f(2\pi) - f(0) = f'(t)2\pi,$$

since the LHS is $(0,0)$, while $f'(t) = (-\sin t, \cos t) \neq (0,0)$.

On the other hand, the mean value theorem of calculus does have a generalization to real–valued functions on a general normed linear space.

Theorem 3.2.3 (Mean Value Theorem I). *Let $f : D \subseteq X \to \mathbb{R}$, where D is an open set in a normed linear space X. Let $a, b \in D$. Assume that the line segment*

$$[a, b] = \{a + t(b - a) : 0 \leq t \leq 1\}$$

is contained in D. If f is continuous on $[a, b]$ and differentiable on the open line segment (a, b), then for some $\xi \in (a, b)$, we have

$$f(b) - f(a) = f'(\xi)(b - a).$$

Proof. Put $g(t) := f(a+t(b-a))$. Then g is continuous on the interval $[0, 1]$ and differentiable on $(0, 1)$. By the Chain Rule,

$$g'(t) = f'(a + t(b - a))(b - a).$$

By the mean value theorem of calculus, there exists $\tau \in (0, 1)$ such that $g(1) - g(0) = g'(\tau)$. Thus

$$\begin{aligned} f(b) - f(a) &= g(1) - g(0) = g'(\tau) = f'(a + \tau(b - a))(b - a) \\ &= f'(\xi)(b - a), \end{aligned}$$

for some $\xi = a + \tau(b - a) \in (a, b)$. □

Theorem 3.2.4 (Mean Value Theorem II). *Let $f : [a, b] \to Y$ be continuous, where Y is a normed linear space. If, for every $x \in (a, b)$, $f'(x)$ exists and satisfies $\|f'(x)\| \leq M$, then $\|f(b) - f(a)\| \leq M(b - a)$.*

Proof. It suffices to prove that for all $\alpha, \beta \in (a, b)$ such that $a < \alpha < \beta < b$, $\|f(\beta) - f(\alpha)\| \leq M(b - a)$, for then the conclusion follows by continuity. Further, it suffices to show that for a fixed $\epsilon > 0$, we have

$$\|f(\beta) - f(\alpha)\| \leq (M + \epsilon)(b - a).$$

Let

$$S := \{x \in [\alpha, \beta] : \|f(x) - f(\alpha)\| \leq (M + \epsilon)(b - a)\}.$$

Note that the set $\{f(x) \in Y : \|f(x) - f(a)\| \leq (M + \epsilon)(b - a)\}$ is a closed ball in Y. By continuity, S is closed. Put $x_0 := \sup S$. Since S is bounded, S is compact, and thus $x_0 \in S$. To complete the proof, it remains to show that $x_0 = \beta$.

By contradiction, suppose that $x_0 < \beta$. Since f is differentiable at x_0, there exists $\delta > 0$ such that $\delta < \beta - x_0$ and for all $|h| < \delta$, we have

$$\|f(x_0 + h) - f(x_0) - f'(x_0)h\| < \epsilon|h|.$$

Put $h := \delta/2$ and take $u := x_0 + \delta/2$. Then $u \in (\alpha, \beta)$, so that

$$\|f(u) - f(x_0) - f'(x_0)(u - x_0)\| < \epsilon(u - x_0)\|.$$

Hence,

$$\|f(u) - f(x_0)\| < \|f'(x_0)(u - x_0)\| + \epsilon(u - x_0) \leq (M + \epsilon)(u - x_0).$$

Since $x_0 \in S$, we have also

$$\|f(x_0) - f(\alpha)\| \leq (M + \epsilon)(x_0 - a).$$

Hence

$$\|f(u) - f(\alpha)\| \leq \|f(u) - f(x_0)\| + \|f(x_0) - f(\alpha)\| \leq (M + \epsilon)(u - a).$$

This shows $u \in S$. Since $u > x_0$, we have a contradiction. Thus $x_0 + \beta$, $\beta \in S$, and

$$\|f(\beta) - f(\alpha)\| \leq (M + \epsilon)(\beta - \alpha) < (M + \epsilon)(b - a).$$

□

Theorem 3.2.5 (Mean Value Theorem III). *Let $f : D \subseteq X \to Y$, where X and Y are normed linear spaces and D is an open subset of X. If the line segment*

$$S := \{ta + (1 - t)b : 0 \leq t \leq 1\}$$

lies in D and if $f'(x)$ exists throughout S, then

$$\|f(b) - f(a)\| \leq \|b - a\| \sup_{x \in S} \|f'(x)\|.$$

Proof. Define $g(t) := f(ta + (1 - t)b)$ for $0 \leq t \leq 1$. By the chain rule, g' exists at $g'(t) = f'(ta + (1 - t)b)(a - b)$. By the second mean value theorem (3.2.4),

$$\|f(b) - f(a)\| = \|g(1) - g(0)\| \leq \sup_{t \in [0,1]} \|g'(t)\| \leq \|b - a\| \sup_{x \in S} \|f'(x)\|.$$

□

Theorem 3.2.6. *Let X and Y be normed linear spaces, let $D \subseteq X$ be a connect open subset of X, and suppose that $f : D \to Y$ is a differentiable map. If $f'(x) = 0$ for all $x \in D$, then f is a constant function.*

Proof. Since $f'(x)$ exists for all $x \in D$, f is continuous throughout D. Choose $x_0 \in D$ and define $A := \{x \in D : f(x) = f(x_0)\}$. By continuity, note that A is closed.

We show that A is also open. Since A is connected, this suffices. Let $x \in A$. Then there exists $r > 0$ so small such that $B(x, r) \subseteq D$, since D is open. If $y \in B(x, r)$, then the line segment $\{tx + (1 - t)y : 0 \leq t \leq 1\}$ lies in $B(x, r)$. By the mean value theorem III (3.2.5),

$$\|f(x) - f(y)\| \leq \|x - y\| \sup_{t \in [0,1]} \|f'(tx + (1 - t)y)\| = 0.$$

Thus $f(y) = f(x) = f(x_0)$, so that $y \in A$ and $B(x, r) \subseteq A$. Hence, A is open, and we have $A = D$. □

3.3. **Extremum Problems and Lagrange Multipliers.**

Definition 3.3.1 (Minimum Point). *A* **minimum point** *of a real–valued function f defined on a set M is a point $x_0 \in M$ such that $f(x_0) \leq f(x)$ for all $x \in M$.*

Definition 3.3.2 (Relative Minimum Point). *Let M be a set and $f : M \to \mathbb{R}$ a real–valued function. A* **relative minimum** *point of f is a point $x_0 \in M$ such that for some neighborhood $\mathcal{N}$ of x_0, we have $f(x_0) \leq f(x)$ for all $x \in \mathcal{N}$.*

Theorem 3.3.3 (Necessary Condition for Extremum). *Let X be a normed linear space, and let $\Omega \subseteq X$ be an open set. Let $f : \Omega \to \mathbb{R}$. If $x_0 \in \Omega$ is a minimum point of f and if $f'(x_0)$ exists, then $f'(x_0) = 0$.*

Proof. By contradiction, suppose that $f'(x_0) \neq 0$. By linearity of $f'(x_0)$, there exists $v \in X$ such that $f'(x_0)v = -1$. Choose $\lambda > 0$ so small that $x_0 + \lambda v \in \Omega$ and

$$\frac{|f(x_0 + \lambda v) - f(x_0) - \lambda f'(x_0)v|}{\lambda\|v\|} < \frac{1}{2\|v\|}.$$

Since $f'(x_0)v = -1$, this gives

$$\left|\frac{1}{\lambda}[f(x_0 + \lambda v) - f(x_0)] - (-1)\right| < \frac{1}{2}.$$

Thus $\frac{1}{\lambda}[f(x_0 + \lambda v) - f(x_0)]$ is within distance $\frac{1}{2}$ of -1, so that

$$\frac{1}{\lambda}[f(x_0 + \lambda v) - f(x_0)] < 0.$$

But this implies

$$f(x_0 + \lambda v) < f(x_0),$$

a contradiction to the assumption. $\square$

We are concerned mostly with *constrained* extremum problems. For example, consider two functions $f, g : \mathbb{R}^2 \to \mathbb{R}$. Put $M := \{(x, y) \in \mathbb{R}^2 : g(x, y) = 0\}$. We look for an extremum of f restricted to M. Suppose that the equation $g(x, y) = 0$ defines y as a function of x, say $y = y(x)$. Then we can look for an *unrestricted* extremum problem of $\phi(x) = f(x, y(x))$. Thus we try to solve the equation $\phi'(x) = 0$. By the Implicit Function Theorem,

$$\begin{aligned} 0 &= f_1(x, y(x)) + f_2(x, y(x))y'(x) \\ &= f_1(x, y(x)) - f_2(x, y(x))g_1(x, y(x))/g_2(x, y(x)). \end{aligned}$$

Thus we must solve simultaneously

$$g(x, y) = 0 \quad \text{and} \quad f_1(x, y) - f_2(x, y)g_1(x, y)/g_2(x, y) = 0.$$

Using the method of Lagrange multipliers, we introduce the function

$$H(x, y, \lambda) := f(x, y) + \lambda g(x, y).$$

We solve simultaneously the equations

$$\begin{cases} \dfrac{\partial H}{\partial x} = 0, \\ \dfrac{\partial H}{\partial y} = 0, \\ \dfrac{\partial H}{\partial}\lambda = 0 \end{cases} \qquad = \qquad \begin{cases} \dfrac{\partial f}{\partial x} + \lambda\dfrac{\partial g}{\partial x} = 0, \\ \dfrac{\partial f}{\partial y} + \lambda\dfrac{\partial g}{\partial y} = 0, \\ g(x, y) = 0 \end{cases} .$$

Example 3.3.4. *Let $f, g : \mathbb{R}^2 \to \mathbb{R}$ be defined by*

$$f(x, y) = x^2 + y^2$$

and

$$g(x, y) = x - y + 1.$$

Note that the set $M := \{(x, y) : g(x, y) = 0\}$ is a straight line in $\mathbb{R}^3$. We note

$$H(x, y, \lambda) = x^2 + y^2 + \lambda(x - y + 1),$$

and the three equations to be solved are

$$\begin{cases} 2x + \lambda = 0, \\ 2y - \lambda = 0, \\ x - y + 1 = 0. \end{cases}$$

The solution is $(x, y) = (-1/2, 1/2)$.

Example 3.3.5. *We find the minimum distance from a point* c *to a line in* $\mathbb{R}^3$. *Let the line be given as the intersection of two planes with equations* $\langle a, x\rangle = k$ *and* $\langle b, x\rangle = \ell$, *where* $a, b, x \in \mathbb{R}^3$. *We take the function* H *to be*

$$H(x_1, x_2, x_3, \lambda, \mu) := \|x - c\|^2 + \lambda[\langle a, x\rangle - k] + \mu[\langle b, x\rangle - \ell].$$

Note that H *is a function of* $(x_1, x_2, x_3, \lambda, \mu)$. *We solve the following system of equations:*

$$\begin{cases} \dfrac{\partial H}{\partial x_1} = 2(x_1 - c_1) + \lambda a_1 + \mu b_1 = 0, \\ \dfrac{\partial H}{\partial x_2} = 2(x_2 - c_2) + \lambda a_2 + \mu b_2 = 0, \\ \dfrac{\partial H}{\partial x_3} = 2(x_3 - c_3) + \lambda a_3 + \mu b_3 = 0, \\ \dfrac{\partial H}{\partial \lambda} = \langle a, x\rangle - k = 0, \\ \dfrac{\partial H}{\partial \mu} = \langle b, x\rangle - \ell = 0. \end{cases}$$

Theorem 3.3.6 (Lagrange Multiplier)**.** *Let* X *be a Banach space and let* $\Omega \subseteq X$ *be an open subset of* X. *Let* $f, g \in \mathcal{C}(\Omega, \mathbb{R})$, *and let* $M := \{x \in \Omega : g(x) = 0\}$. *If* x_0 *is a local minimum point of* $f|_M$ *and if* $g'(x_0) \neq 0$, *then* $f'(x_0) = \lambda g'(x_0)$ *for some* $\lambda \in \mathbb{R}$.

Proof. Let U be a neighborhood of x_0 such that for all $x \in U \cap M$, we have $f(x_0) \leq f(x)$. We may assume that $U \subseteq \Omega$, for otherwise define $U := U \cap \Omega$. Define a function $F : U \to \mathbb{R}^2$ by

$$F(x) := (f(x), g(x)).$$

Then notice that

$$F(x_0) = (f(x_0), g(x_0)) = (f(x_0), 0),$$

and

$$F'(x)v = (f'(x)v, g'(x)v)$$

for all $v \in X$. Observe that if $r < f(x_0)$, then $(r, 0)$ is not in $F(U)$, for otherwise we get a contradiction to the assumption that $f'(x_0)$ is a minimum point. Thus $F(U)$ is not a neighborhood of $F(x_0)$. Hence, $F'(x_0)$ is not surjective, as a linear map from X to $\mathbb{R}^2$. It follows

$$F'(x_0)v = \alpha(v)(\theta, \mu)$$

for some continuous linear functional $\alpha \in X^*$. Then we have that $f'(x_0)v = \alpha(v)\theta$ and $g'(x_0)v = \alpha(v)\mu$. Since $g'(x_0) \neq 0$, $\mu \neq 0$. Therefore,

$$f'(x_0)v = \frac{\theta}{\mu}\alpha(v)\mu = \frac{\theta}{\mu}g'(x_0)v.$$

This is for all $v \in X$, so that this completes the proof. $\square$

Theorem 3.3.7 (Lagrange Multipliers). *Let X be a Banach space and let $\Omega \subseteq X$ be an open subset. Let $f, g_1, \ldots, g_n \in \mathcal{C}(\Omega, \mathbb{R})$ and let $M := \{x \in \Omega : g_1(x) = \cdots = g_n(x) = 0\}$. If x_0 is a local minimum point of $f|_M$, then there exist $\mu, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$ not all zero such that*

$$\mu f'(x_0) + \lambda_1 g_1'(x_0) + \lambda_2 g_2'(x_0) + \cdots + \lambda_n g_n'(x_0) = 0.$$

Proof. Let U be a neighborhood of x_0 such that $U \subseteq \Omega$ and so that $f(x_0) \leq f(x)$ for all $x \in U \cap M$. Define $F : U \to \mathbb{R}^{n+1}$ by

$$F(x) := (f(x), g_1(x), g_2(x), \ldots, g_n(x)).$$

If $r < f(x_0)$, then the point $(r, 0, 0, \ldots, 0)$ is not in $F(U)$. Thus $F(U)$ does not contain a neighborhood of the point

$$(f(x_0), g_1(x_0), \ldots, g_n(x_0)) = (f(x_0), 0, 0 \ldots, 0).$$

It follows that $F'(x_0)$ is not surjective. Since the range of $F'(x_0)$ is a linear subspace of $\mathbb{R}^{n+1}$, we see that it is a proper subspace of $\mathbb{R}^{n+1}$. Thus it is contained in a hyperplane through the origin. This means that for some $\mu, \lambda_1, \ldots, \lambda_n \in \mathbb{R}$, not all zero, we have

$$\mu f'(x_0)v + \lambda_1 g_1'(x_0)v + \cdots + \lambda_n g_n'(x_0)v = 0$$

for all $v \in X$. This completes the proof. $\square$

Example 3.3.8. *Let X be a Hilbert space and let $A : X \to X$ be a compact operator. Then $\|A\| = \max\{|\lambda| : \lambda \in \Lambda(A)\}$, where $\Lambda(A)$ denotes the set of eigenvalues of A. Then recall that*

$$\|A\| = \sup\{|\langle Ax, x\rangle| : \|x\| = 1\}.$$

Thus we can find an eigenvalue of A by determining an extremum of $\langle Ax, x\rangle$ on the set defined by $\|x\| = 1$.

Lemma 3.3.9. *If A is Hermitian, then the Rayleigh quotient*

$$f(x) := \frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$

has a stationary value at each eigenvector.

Proof. Let A be Hermitian, and suppose that $Ax = \lambda x$, where $x \neq 0$. Then $f(x) = \langle Ax, x\rangle / \langle x, x\rangle = \langle \lambda x, x\rangle / \langle x, x\rangle = \lambda$. Recall that the eigenvalues of a Hermitian operator are real. Taking the derivative of f, we find

$$\begin{aligned}
\lim_{h\to 0} \frac{|f(x+h) - f(x)|}{\|h\|} &= \lim_{h\to 0} \frac{\left|\frac{\langle Ax+Ah, x+h\rangle}{\langle x+h, x+h\rangle} - \lambda\right|}{\|h\|} \\
&= \lim_{h\to 0} \frac{|\langle Ax, x\rangle + \langle Ah, x\rangle + \langle Ax, h\rangle + \langle Ah, h\rangle - \lambda\|x+h\|^2|}{\|h\|\|x+h\|^2} \\
&= \lim_{h\to 0} \frac{|\langle h, Ax\rangle + \lambda\langle x, h\rangle + \langle Ah, h\rangle - 2\lambda \mathrm{Re}\langle x, h\rangle - \lambda\langle h, h\rangle|}{\|h\|\|x+h\|^2} \\
&= \lim_{h\to 0} \frac{|\lambda\langle h, x\rangle + \lambda\langle x, h\rangle + \langle Ah, h\rangle - 2\lambda \mathrm{Re}\langle x, h\rangle - \lambda\langle h, h\rangle|}{\|h\|\|x+h\|^2}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{h\to 0} \frac{|\langle Ah, h\rangle - \lambda \langle h, h\rangle|}{\|h\|\|x+h\|^2} \\
&= \lim_{h\to 0} \frac{|\langle Ah - \lambda h, h\rangle|}{\|h\|\|x+h\|^2} \\
&\leq \lim_{h\to 0} \frac{\|Ah - \lambda h\|\|h\|}{\|h\|\|x+h\|^2} \\
&\leq \lim_{h\to 0} \frac{\|A - \lambda I\|\|h\|}{\|x+h\|^2} = 0.
\end{aligned}
$$

Thus $f'(x) = 0$. □

Notice that we defined the Rayleigh quotient

$$\frac{\langle Ax, x\rangle}{\|x\|^2}.$$

We may write this as

$$\frac{\langle Ax, x\rangle}{\|x\|^2} = \left\langle A\left(\frac{x}{\|x\|}\right), \frac{x}{\|x\|}\right\rangle.$$

Thus it is possible to consider the simpler function $F(x) = \langle Ax, x\rangle$ restricted to the unit sphere.

Theorem 3.3.10. *Let X be a Hilbert space and let $A : X \to X$ be a Hermitian operator. Then each local constrained minimum or maximum point of $\langle Ax, x\rangle$ on the unit sphere is an eigenvector of A. Moreover, the value of $\langle Ax, x\rangle$ is the corresponding eigenvalue.*

Proof. Put $F(x) := \langle Ax, x\rangle$ and $G(x) := \|x\|^2 - 1$. Then

$$F'(x)h = 2\mathrm{Re}\,\langle Ax, h\rangle \quad \text{and} \quad G'(x)h = 2\mathrm{Re}\,\langle x, h\rangle\,.$$

Let $x \in X$ be an extremum point of F on the set $M := \{x \in X : G(x) = 0\}$. By the method of Lagrange multipliers in Theorem (3.3.6), there exist $\mu, \lambda \in \mathbb{R}$, not both zero, such that

$$\mu F'(x) + \lambda G'(x) = 0.$$

Since $\|x\| = 1$, $G'(x) \neq 0$, and thus $\mu \neq 0$. By homogeneity, we set $\mu := -1$. Thus

$$
\begin{aligned}
0 &= -2\mathrm{Re}\,\langle Ax, h\rangle + 2\lambda\mathrm{Re}\,\langle x, h\rangle = -2\mathrm{Re}\,\langle Ax, h\rangle + 2\mathrm{Re}\,\langle \lambda x, h\rangle \\
&= 2\mathrm{Re}\,\langle \lambda x - Ax, h\rangle = -2\mathrm{Re}\,\langle (A - \lambda I)x, h\rangle
\end{aligned}
$$

for all $h \in X$. Hence, $Ax = \lambda x$. □

Extremum problems with inequality constraints can also be discussed, leading to Kuhn–Tucker Theory.

Definition 3.3.11 (Ordered Vector Space)**.** *An* **ordered vector space** *is a pair $(X, \geq)$ in which X is a real vector space and $\geq$ is a partial order on X that is consistent with the linear structure, that is,*

$$x \geq y \implies x + z \geq y + z,$$

and

$$x \geq y, \lambda \geq 0 \implies \lambda x \geq \lambda y.$$

If $(X, \geq)$ is an ordered vector space, then X^* is ordered in a standard way; namely, we define $\phi \geq 0$ to mean $\phi(x) \geq 0$ for all $x \geq 0$.

Consider for example the space $\mathcal{C}[a,b]$, in which the order $f \geq g$ is defined to mean $f(t) \geq g(t)$ for all $t \in [a,b]$. The conjugate space consists of signed measures.

In the following theorem, we seek necessary conditions for a point x_0 to maximize $f(x)$ subject to $G(x) \geq 0$.

Theorem 3.3.12. *Let X be a normed linear space and let $(Y, \geq)$ be an ordered vector space. Let $f : X \to \mathbb{R}$ and $G : X \to Y$ be differentiable. If x_0 is a local maximum point of f on the set $\{x \in X : G(x) \geq 0\}$ and if there is an $h \in X$ such that $G(x_0) + G'(x_0)h$ is an interior point of the positive cone $K := \{x \in X : x \geq 0\}$, then there exists a nonnegative functional $\phi \in Y^*$ such that $\phi(G(x_0)) = 0$ and $f'(x_0) = -\phi \circ G'(x_0)$.*

3.4. **Calculus of Variations.** The calculus of variations, interpreted broadly, deals with extremum problems involving *functions*. It is analogous to the theory of maxima and minima in elementary calculus, but with the added complication that the unknowns in the problems are not simple numbers – they are functions.

Example 3.4.1. *Find the equation of an arc of minimal length joining two points in the place. Let the points be (a, α) and (b, β), where $a < b$. Let the arc be given by a continuously differentiable function $y = y(x)$, where $y(a) = \alpha$ and $y(b) = \beta$. The arc length is given by the integral*

$$\int_a^b \sqrt{1 + y'(x)^2}\ dx.$$

The solution is obviously a straight line, which will be proved later.

Example 3.4.2. *We find a function $y \in \mathcal{C}^1[a,b]$, satisfying $y(a) = \alpha$ and $y(b) = \beta$, such that the surface of revolution obtained by rotating the graph of y about the $x-$axis has minimum area. To solve this, recall from calculus that the area to be minimized is given by*

$$\int_a^b 2\pi y(x)\ ds = 2\pi \int_a^b y(x)\sqrt{1 + y'(x)^2}\ dx.$$

Notice that the above examples have a common form, that is, in each one, there is a nonlinear functional of the form

$$\int_a^b F(x, y(x), y'(x))\ dx$$

to be minimized. The unknown function y is required to satisfy endpoint conditions $y(a) = \alpha$ and $y(b) = \beta$. In addition, some smoothness conditions must be imposed on y, since the functional is allowed to involve y'. The first theorem establishes a necessary condition for the extrema, known as the Euler–Lagrange Equation.

Lemma 3.4.3. *If $v : [a,b] \to \mathbb{R}$ is piecewise continuous and if $\int_a^b u(x)v(x)\ dx = 0$ for every $u \in \mathcal{C}^1[a,b]$ that vanishes at the endpoints a and b, then $v \equiv 0$.*

Proof. By contradiction, suppose that $v \neq 0$. Then there exists a nonempty open interval $(\alpha, \beta) \subseteq [a,b]$ such that v is continuous on (α, β) and does not vanish for any $t \in (\alpha, \beta)$.

Without loss of generality, we assume that $v(t) > 0$ throughout (α, β). There exists a function $u \in \mathbb{C}^1[a, b]$ such that $u(t) > 0$ throughout (α, β) and $u(t) = 0$ elsewhere in $[a, b]$. Thus

$$\int_a^b u(x)v(x)\ dx = \int_\alpha^\beta u(x)v(x)\ dx > 0,$$

a contradiction. $\square$

Theorem 3.4.4 (The Euler–Lagrange Equation)**.** *Let $F : \mathbb{R}^3 \to \mathbb{R}$ be a function with piecewise continuous second order partial derivatives. Suppose that $y \in \mathcal{C}^1[a, b]$ minimizes the nonlinear functional*

$$\begin{cases} \displaystyle\int_a^b F(x, y(x), y'(x))\ dx, \\ y(a) = \alpha, \quad y(b) = \beta. \end{cases}$$

Then the Euler–Lagrange Equation

$$\frac{d}{dx} F_3(x, y(x), y'(x)) = F_2(x, y(x), y'(x))$$

is satisfied.

Proof. Let $u \in \mathcal{C}^1[a, b]$ be such that $u(a) = u(b) = 0$. Suppose that $y \in \mathcal{C}^1[a, b]$ is a solution of the problem. Notice that the function $\widetilde{y} := y + tu \in \mathcal{C}^1[a, b]$ satisfies $\widetilde{y}(a) = \alpha$ and $\widetilde{y}(b) = \beta$. We find

$$\begin{aligned} 0 &= \frac{d}{dt} \int_a^b F(x, y(x) + tu(x), y'(x) + tu'(x))\ dx \bigg|_{t=0} \\ &= \int_a^b [F_2(x, y(x) + tu(x), y'(x) + tu'(x))u(x) + F_3(x, y(x) + tu(x), y'(x) + tu'(x))u'(x)] \bigg|_{t=0} dx \\ &= \int_a^b F_2(x, y(x), y'(x))u(x) + F_3(x, y(x), y'(x))u'(x)\ dx \end{aligned}$$

Integrating $\int_a^b F_3(x, y(x), y'(x))u'(x)\ dx$ by parts and noting that u vanishes at the endpoints, we obtain

$$\begin{aligned} 0 &= \int_a^b F_2(x, y(x), y'(x))u(x)\ dx + u(x)F_3(x, y(x), y'(x))|_a^b - \int_a^b \frac{d}{dx} \left[F_3(x, y(x), y'(x))\right] u(x)\ dx \\ &= \int_a^b F_2(x, y(x), y'(x))u(x) - \frac{d}{dx} \left[F_3(x, y(x), y'(x))\right] u(x)\ dx \\ &= \int_a^b \left[F_2(x, y(x), y'(x)) - \frac{d}{dx} F_3(x, y(x), y'(x))\right] u(x)\ dx. \end{aligned}$$

Since u was arbitrary, we have by the preceding Lemma (3.4.3) that

$$F_2(x, y(x), y'(x)) - \frac{d}{dx} F_3(x, y(x), y'(x)) = 0,$$

which completes the proof. $\square$

Example 3.4.5. *Recall the Example* (3.4.2)*. In this problem,*

$$F(u, v, w) = \sqrt{1 + w^2}.$$

We note that $F_1 \equiv 0, F_2 \equiv 0,$ *and*

$$F_3 = w(1+w^2)^{-1/2}.$$

Hence,

$$F_3(x, y(x), y'(x)) = y'(x)[1 + y'(x)^2]^{-1/2}.$$

Thus the Euler–Lagrange Equation gives

$$\frac{d}{dx} F_3(x, y(x), y'(x)) = 0.$$

Integrating, we have

$$F_3(x, y(x), y'(x)) = C,$$

for some $C \in \mathbb{R}$*. Solving for* $y'(x)$*, we find that* $y'(x)$ *must be constant. Thus* $y(x) = \alpha + m(x - a)$*, where* $m = (\beta - \alpha)/(b - a)$*.*

Theorem 3.4.6. *Assume the hypotheses of Theorem* (3.4.4)*. If, in addition,* $F_1 \equiv 0$*, then the Euler–Lagrange Equation implies that*

$$y'(x)F_3(x, y(x), y'(x)) - F(x, y(x), y'(x)) = C,$$

for some $C \in \mathbb{R}$*.*

Proof. We show that $\frac{d}{dx}[y'(x)F_3(x, y(x), y'(x)) - F(x, y(x), y'(x))] = 0$. Observe, by the Chain Rule and the Euler–Lagrange Equation,

$$\begin{aligned}
&\frac{d}{dx}\left[y'(x)F_3(x, y(x), y'(x)) - F(x, y(x), y'(x))\right] \\
&\quad = y''(x)F_3(x, y(x), y'(x)) + y'(x)\frac{d}{dx}[F_3(x, y(x), y'(x))] - F_1(x, y(x), y'(x)) - \\
&\qquad\quad y'(x)F_2(x, y(x), y'(x)) - y''(x)F_3(x, y(x), y'(x)) \\
&\quad = y'(x)\frac{d}{dx}[F_3(x, y(x), y'(x)) - F_2(x, y(x), y'(x))] \\
&\quad = 0,
\end{aligned}$$

which completes the proof. □

Theorem 3.4.7. *Any function* $y \in \mathbb{C}^2[a, b]$ *that minimizes the integral equation*

$$\int_a^b F(x, y(x), y'(x))\, dx$$

subject to the endpoint constraint $y(a) = \alpha$ *must satisfy the two conditions*

$$\frac{d}{dx} F_3(x, y(x), y'(x)) = F_2(x, y(x), y'(x))$$

and

$$F_3(b, y(b), y'(b)) = 0.$$

4. Basic Approximate Methods in Analysis

4.1. The Method of Iteration.

Definition 4.1.1 (Iteration)**.** *Let X be a normed linear space, and let $F : X \to X$ be continuous. Fix $x_0 \in X$. By* **iteration***, we mean one of the following two processes:*

$$x_{n+1} = Fx_n,\ n \in \mathbb{N}_0, \qquad \text{or} \qquad x_n = F^n x_0,\ n \in \mathbb{N}_0.$$

We see that if $\lim_{n\to\infty} x_n$ exists, then it is a fixed point of F, for

$$F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} Fx_n = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} x_n.$$

Thus we see that the method of iteration is one technique for finding fixed points of operators.

Definition 4.1.2 (Contraction Mapping)**.** *Let (X, d) be a metric space. A mapping F is called a* **contraction mapping** *if there exists $0 < \theta < 1$ such that for all $x, y \in X$,*

$$d(Fx, Fy) < \theta d(x, y).$$

Recall that every Banach space is necessarily a complete metric space with the induced metric $d(x, y) = \|x - y\|$. Moreover, a closed set in a Banach space is also a complete metric space.

The following theorem is due to Banach, 1922.

Theorem 4.1.3 (Contraction Mapping Theorem)**.** *Let (X, d) be a complete metric space. If F is a contraction mapping on X, then there exists a unique fixed point $\xi \in X$. Moreover, for any $x \in X$, the point ξ is the limit of every sequence $\{F^n x\}_{n\in\mathbb{N}_0}$.*

Proof. Choose $x_0 \in X$. Since F is a contraction mapping, there exists $0 < \theta < 1$ such that

$$d(F^n x_0, F^{n-1}x_0) = d(F(F^{n-1}x_0), F(F^{n-2}x_0)) < \theta d(F^{n-1}x_0, F^{n-2}x_0).$$

Repeating this process $n - 2$ subsequent times, we find

$$d(F^n x_0, F^{n-1}x_0) \le \theta^{n-1} d(Fx_0, x_0).$$

We show that $\{F^n x_0\}_{n\in\mathbb{N}_0}$ is Cauchy. Let $m, n > N$ be such that $m \ge n$. Then

$$\begin{aligned} d(F^m x_0, F^n x_0) &\le d(F^m x_0, F^{m-1}x_0) + d(F^{m-1}x_0, F^{m-2}x_0) + \cdots + d(F^{n+1}x_0, F^n x_0) \\ &\le [\theta^{m-1} + \theta^{m-2} + \cdots + \theta^n] d(Fx_0, x_0) \\ &\le \left\{ \sum_{n=N}^{\infty} \theta^n \right\} d(Fx_0, x_0) \\ &\le \frac{\theta^N}{1-\theta} d(Fx_0, x_0). \end{aligned}$$

Since $0 < \theta < 1$, $\lim_{N\to\infty} \theta^N = 0$. Thus the sequence $\{F^n x_0\}_{n\in\mathbb{N}_0}$ is Cauchy. Since the space X is complete, the sequence converges to a point $\xi \in X$. Since the contractive property of F implies the continuity of F, it follows that ξ is a fixed point of F.

We now show uniqueness of ξ. Suppose that $\eta \in X$ is another fixed point of F. Then we have

$$d(\xi, \eta) = d(F\xi, F\eta) \le \theta d(\xi, \eta).$$

If $\xi \ne \eta$, then $d(\xi, \eta) > 0$, and thus the above inequality implies that $\theta \ge 1$, a contradiction. Hence, the fixed point ξ is unique. $\square$

We illustrate the iterative process with a Fredholm integral equation.

Definition 4.1.4 (Fredholm Integral Equation of the Second Kind)**.** *Let $w \in \mathcal{C}[0,1]$ and $K \in \mathcal{C}([0,1]\times[0,1]\times\mathbb{R},\mathbb{R})$. A* **Fredholm integral equation of the second kind** *is defined by $x = Fx$, where*

$$(Fx)(t) := \int_a^b K(s,t,x(s))\ ds + w(t).$$

We determine a solution $x \in \mathcal{C}[0,1]$. Recall that the space $(\mathcal{C}[0,1], \|\cdot\|_\infty)$ is complete. To see whether F is a contraction, we calculate $\|Fu - Fv\|$:

$$\begin{aligned}\|Fu - Fv\| &= \sup_{t\in[0,1]} \left|\int_0^1 K(s,t,u(s)) - K(s,t,v(s))\ ds\right| \\ &\leq \sup_{t\in[0,1]} \int_0^1 |K(s,t,u(s)) - K(s,t,v(s))|\ ds.\end{aligned}$$

Thus if K satisfies a Lipschitz condition in the third argument

$$|K(s,t,\xi) - K(s,t,\eta)| \leq \theta|\xi-\eta|,$$

where $0 < \theta < 1$, then we obtain

$$\|Fu - Fv\| \leq \sup_{t\in[0,1]} \int_0^1 \theta|u(s)-v(s)|\ ds \leq \int_0^1 \theta\|u-v\| = \theta\|u-v\|.$$

Hence, by the Banach Fixed Point Theorem, the iteration $x_n = F^n x_0$ leads to a solution, starting from any $x_0 \in \mathcal{C}[0,1]$.

We get the following theorem.

Theorem 4.1.5. *Let $K \in \mathcal{C}([0,1]\times[0,1]\times\mathbb{R},\mathbb{R})$ satisfy a Lipschitz condition in the third argument*

$$|K(s,t,\xi) - K(s,t,\eta)| \leq \theta|\xi-\eta|,$$

with $0 < \theta < 1$. Then the Fredholm integral equation

$$x = (Fx)(t) = \int_0^1 K(s,t,x(s))\ ds + w(t)$$

has a unique solution in $\mathcal{C}[0,1]$.

Example 4.1.6. *Consider the nonlinear Fredholm equation*

$$x(t) = \frac{1}{2}\int_0^1 \cos(stx(s))\ ds.$$

Define $K(s,t,\xi) := \frac{1}{2}\cos(st\xi)$. We see that, by the Mean Value Theorem,

$$|K(s,t,\xi) - K(s,t,\eta)| = \frac{1}{2}|\cos(st\xi) - \cos(st\eta)| = \frac{1}{2}|\sin(st\zeta)|\xi-\eta| \leq \frac{1}{2}|\xi-\eta|.$$

Thus Theorem (4.1.5) is applicable with $\theta = \frac{1}{2}$.

Say the iteration is begun with $x_0 := 0$. Then

$$x_1(t) = (Fx_0)(t) = \frac{1}{2}\int_0^1\ ds = \frac{1}{2},$$

and

$$x_2(t) = (Fx_1)(t) = \frac{1}{2}\int_0^1 \cos\left(\frac{1}{2}st\right)\ ds = \frac{1}{t}\sin\left(\frac{1}{2}t\right).$$

The method of iteration can also be applied to differential equations, usually by first turning them into equivalent integral equations. This procedure is of high importance, as it is capable of yielding existence theorems with very little effort.

Theorem 4.1.7. *Let $S := [0, b]$, where $b > 0$. Let a continuous map $f : S \times \mathbb{R} \to \mathbb{R}$ satisfy a Lipschitz condition in the second argument*

$$|f(s, t_1) - f(s, t_2)| \leq \lambda |t_1 - t_2|,$$

where $\lambda > 0$ is a constant depending only on f. Then the initial value problem

$$x'(s) = f(s, x(s)), \quad x(0) = \beta \tag{4.1.0.1}$$

has a unique solution in $\mathcal{C}(S)$.

Proof. We first show that the initial value problem (4.1.0.1) is equivalent to the integral equation

$$x = (Ax)(s) := \beta + \int_0^s f(t, x(t))\ dt. \tag{4.1.0.2}$$

First suppose that $x \in \mathcal{C}(S)$ solves (4.1.0.1). Then it follows

$$\begin{aligned}(Ax)(s) &= \beta + \int_0^s f(t, x(t))\ dt = \beta + \int_0^s x'(t)\ dt = \beta + x(s) - x(0) \\ &= x(s),\end{aligned}$$

so that x solves (4.1.0.2). Now assume that x solves (4.1.0.2). Then

$$\begin{aligned}x'(s) &= \frac{d}{ds}(Ax)(s) = \frac{d}{ds}\left\{\beta + \int_0^s f(t, x(t))\ dt\right\} = f(s, x(s)) + \int_0^s \frac{\partial}{\partial s} f(t, x(t))\ dt \\ &= f(s, x(s)).\end{aligned}$$

Moreover,

$$x(0) = (Ax)(0) = \beta + \int_0^0 f(t, x(t))\ dt = \beta,$$

so that x solves (4.1.0.2).

We now introduce a new norm on $\mathcal{C}(S)$ by defining

$$\|x\|_w := \sup_{s \in S} |x(s)| e^{-2\lambda s}.$$

The space $(\mathcal{C}(S), \|\cdot\|_w)$ is complete. By the equivalence of (4.1.0.1) and (4.1.0.2), it suffices to show that A has a fixed point. We show that A is a contraction. Let $u, v \in (\mathcal{C}(S), \|\cdot\|_w)$. Then, for $s \in [0, b]$, we have

$$\begin{aligned}|(Au - Av)(s)| &= \left|\int_0^s f(t, u(t)) - f(t, v(t))\ dt\right| \leq \int_0^s |f(t, u(t)) - f(t, v(t))|\ dt \\ &\leq \int_0^s \lambda |u(t) - v(t)|\ dt \\ &= \lambda \int_0^s e^{2\lambda t} e^{-2\lambda t} |u(t) - v(t)|\ dt \\ &\leq \lambda \|u - v\|_w \int_0^s e^{2\lambda t}\ dt \\ &= \lambda \|u - v\|_w \left(\frac{e^{2\lambda s}}{2\lambda} - \frac{1}{2\lambda}\right)\end{aligned}$$

$$\leq \lambda\|u-v\|_w \frac{e^{2\lambda s}}{2\lambda}$$
$$= \frac{1}{2}e^{2\lambda s}\|u-v\|_w.$$

Thus
$$e^{-2\lambda s}|(Au-Av)(s)| \leq \frac{1}{2}\|u-v\|_w,$$
so that
$$\|Au-Av\|_w \leq \frac{1}{2}\|u-v\|_w,$$
which shows that A is a contraction. By the Contraction Mapping Theorem (4.1.3), A has a unique fixed point. This completes the proof. □

Example 4.1.8. *Consider the following initial value problem:*
$$\begin{cases} x' = \cos(xe^s), & x \in \mathcal{C}[0,10], \\ x(0) = 0. \end{cases}$$
Note that $f(s,t) = \cos(te^s)$. By the Mean Value Theorem,
$$|f(s,t_1) - f(s,t_2)| = \left|\frac{\partial f}{\partial t}(s,\tau)\right| |t_1 - t_2|.$$
For $s \in [0,10]$ and $t \in \mathbb{R}$,
$$\left|\frac{\partial f}{\partial t}\right| = |-e^s \sin(te^s)| = e^s|\sin(te^s)| \leq e^{10}.$$
Hence f satisfies a Lipschitz condition in the second argument, and the IVP has a unique solution in $\mathcal{C}[0,10]$ by Theorem (4.1.7).

The iteration described above Theorem (4.1.5) is often referred to as *Picard* iteration.

Example 4.1.9. *Consider the following initial value problem:*
$$\begin{cases} x' = 2t(1+x), & x \in \mathcal{C}[0,b], \\ x(0) = 0. \end{cases}$$
The formula for Picard iteration for this IVP is
$$x_{n+1}(s) := (Ax_n)(s) = \int_0^s 2t(1+x_n(t))\ dt = s^2 + 2\int_0^s tx_n(t)\ dt.$$
If $x_0 := 0$, then
$$x_1(s) = s^2,$$
$$x_2(s) = s^2 + 2\int_0^s t^3\ dt = s^2 + \frac{1}{2}s^4,$$
$$x_3(s) = s^2 + 2\int_0^s t^3 + \frac{1}{2}t^5\ dt = s^2 + \frac{1}{2}s^4 + \frac{1}{6}s^6.$$
We see taht the partial sums tend to $x(s) = e^{s^2} - 1$.

Theorem 4.1.10. *Let (X,d) be a complete metric space, and suppose that $F : X \to X$ is a mapping such that for some $m \in \mathbb{N}_0$, F^m is a contraction. Then F has a unique fixed point $\xi \in X$, and ξ is the limit of every sequence $\{F^n x\}_{n\in\mathbb{N}_0}$, for any $x \in X$.*

Proof. Since F^m is a contraction mapping, F^m has a unique fixed point $\xi \in X$, by Theorem (4.1.3). Thus

$$F\xi = F(F^m\xi) = F^{m+1}\xi = F^m(F\xi).$$

Thus $F\xi$ is also a fixed point of F^m. By uniqueness, $F\xi = \xi$, so that F has at least one fixed point, namely, ξ. If x is any fixed point of F, then

$$Fx = x, \quad F^2x = F(Fx) = Fx = x, \quad \cdots \quad F^m x = F^{m-1}Fx = F^{m-1}x = x.$$

Thus x is a fixed point of F^m, and by uniqueness, $x = \xi$. That is, ξ is the only fixed point of F.

It remains to show that the sequence $F^n x$ converges to ξ as $n \to \infty$. For $i \in \{1, 2, \ldots, m\}$, we have

$$F^{nm+i}x = F^{nm}(F^i x) \to \xi$$

as $n \to \infty$. Fix $\epsilon > 0$. Since ξ is a fixed point of F^m, there exists a positive integer N so large that for all $n \geq N$, we have $d(F^{nm+i}x, \xi) < \epsilon$. Since each $j \in \mathbb{N}$ such that Nm can be written as $j := nm + i$, where $n \geq N$ and $i \in \{1, 2, \ldots, m\}$ by the division algorithm, we have $d(F^j x, \xi) < \epsilon$ for all $j \geq Nm$. This proves that $\lim_{j\to\infty} F^j x = \xi$, which completes the proof. $\square$

Definition 4.1.11 (Linear Volterra Equation of the Second Kind)**.** *Let $v \in \mathcal{C}[a, b]$ and $K \in \mathcal{C}([a, b] \times [a, b], \mathbb{R})$. A* **linear Volterra equation of the second kind** *is defined by*

$$x(t) = v(t) + \int_a^t K(t, s)x(s)\ ds, \quad x \in \mathcal{C}[a, b].$$

Note that a Volterra equation can be written as

$$x = Ax + v,$$

where

$$(Ax)(t) := \int_0^t K(t, s)x(s)\ ds.$$

We see that

$$|(Ax)(t)| = \left|\int_a^t K(t, s)x(s)\ ds\right| \leq \int_a^t |K(t, s)||x(s)|\ ds \leq \|K\|_\infty \|x\|_\infty (t - a).$$

From this it follows that

$$\begin{aligned} |(A^2x)(t)| &= \left|\int_a^t K(t, s)(Ax)(s)\ ds\right| \leq \int_a^t |K(t, s)||(Ax)(s)|\ ds \\ &\leq \int_a^t |K(t, s)|\|K\|_\infty \|x\|_\infty (s - a)\ ds \\ &\leq \|K\|_\infty^2 \|x\|_\infty \int_a^t s - a\ ds \\ &= \|K\|_\infty^2 \|x\|_\infty \frac{(t - a)^2}{2}. \end{aligned}$$

Repetition of this process gives

$$|(A^n x)(t)| \leq \|K\|_\infty^n \|x\|_\infty \frac{(t - a)^n}{n!} \leq \|K\|_\infty^n \|x\|_\infty \frac{(b - a)^n}{n!}.$$

Hence,

$$\|A^n\| = \sup_{\|x\|=1} \|A^n x\| = \frac{(\|K\|_\infty (b - a))^n}{n!}.$$

Choose $m \in \mathbb{N}_0$ so large that $\|A^m\| < 1$. Write

$$(Fx)(t) := v(t) + \int_a^t K(t,s)x(s)\, ds,$$

and note $Fx = Ax + v$. Then

$$F^2x = F(Fx) = F(Ax+v) = A(Ax+v) + v = A^2x + Av + v,$$
$$F^3x = F^2(Fx) = A^2(Fx) + Av + v = A^2(Ax+v) + Av + v = A^3x + A^2v + Av + v,$$

and so on. Thus

$$\|F^mx - F^my\| = \|A^mx - A^my\| \le \|A^m\|\|x-y\| < \|x-y\|,$$

which shows that F^m is a contraction. By Theorem (4.1.10), F has a unique fixed point, which we may obtain through iteration of the map F. Notice also that this conclusion is reached without making strong assumptions regarding the kernel K. In particular, K need only be bounded throughout $[a,b] \times [a,b]$.

Theorem 4.1.12. *Let $v \in \mathcal{C}[a,b]$ and let $K \in \mathcal{C}([a,b]\times[a,b], \mathbb{R})$. Then the integral equation*

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

has a unique solution in $\mathcal{C}[a,b]$.

Proof. The proof follows immediately from the preceding argument. □

Theorem 4.1.13. *Let X be a Hilbert space, and let $F : X \to X$ be a mapping such that for all $x, y \in X$,*

(1) $\langle Fx - Fy, x - y\rangle \ge \alpha\|x-y\|^2$, for some $\alpha > 0$;
(2) $\|Fx - Fy\| \le \beta\|x-y\|$, for some $\beta > 0$.

Then F is bijective. Consequently, F^{-1} exists.

Proof. We first show that F is injective. Suppose that $Fx = Fy$ for some $x, y \in X$. Then, by (1),

$$\alpha\|x-y\|^2 \le \langle Fx - Fy, x - y\rangle = 0.$$

Since $\alpha > 0$, it follows that $x = y$.

To see that F is surjective, let $w \in X$. We show that there exists $x \in X$ such that $Fx = w$. It is equivalent to show that, for any $\lambda > 0$, there exists $x \in X$ satisfying $x - \lambda(Fx - w) = x$. Define

$$Gx := x - \lambda(Fx - w).$$

We show the existence of a fixed point of G. Letting $\lambda = \alpha/\beta^2$, we obtain

$$\begin{aligned}
\|Gx - Gy\|^2 &= \|x - \lambda(Fx - w) - y + \lambda(Fy - w)\|^2 \\
&= \|x - y - \lambda(Fx - Fy)\|^2 \\
&= \|x-y\|^2 - 2\lambda\langle Fx - Fy, x - y\rangle + \lambda^2\|Fx - Fy\|^2 \\
&\le \|x-y\|^2 - 2\lambda\alpha\|x-y\|^2 + \lambda^2\beta^2\|x-y\|^2 \\
&= \|x-y\|^2(1 - 2\lambda\alpha + \lambda^2\beta^2)
\end{aligned}$$

$$= \|x-y\|^2 \left(1 - 2\frac{\alpha^2}{\beta^2} + \frac{\alpha^2}{\beta^2}\right)$$
$$= \|x-y\|^2 \left(1 - \frac{\alpha^2}{\beta^2}\right).$$

Note that $\alpha \leq \beta$, for by the Cauchy–Schwarz Inequality and (2),

$$|\langle Fx - Fy, x - y\rangle| \leq \|Fx - Fy\|\|x - y\| \leq \beta\|x - y\|^2,$$

which implies

$$\beta\|x-y\|^2 \geq \alpha\|x-y\|^2.$$

This completes the proof. □

5. Distributions

5.1. **Definition and Examples.** The objective of distributions is to treat functions and functionals, and to notice that when so interpreted, differentiation is always possible. The functionals that now become the focus of study are called "distributions," or "generalized functions."

Definition 5.1.1 (Multi–Index). *A* **multi–index** *is any $n-$tuple of nonnegative integers*

$$\alpha := (\alpha_1, \alpha_2, \dots, \alpha_n).$$

Definition 5.1.2 (Order of Multi–Index). *The* **order** *of a multi–index α is the quantity*

$$|\alpha| := \sum_{k=1}^{n} \alpha_k.$$

For any multi–index $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$, we associate a partial differential operator D^α corresponding to it. We define

$$D^\alpha := \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1}\left(\frac{\partial}{\partial x_2}\right)^{\alpha_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{\alpha_n} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\dots\partial x_n^{\alpha_n}}.$$

This operates on functions of n real variables $x_1, x_2, \dots, x_n$.

Example 5.1.3. *If $n = 3$ and $\alpha = (3, 0, 4)$, then*

$$D^\alpha\phi = \left(\frac{\partial\phi}{\partial x_1}\right)^3\left(\frac{\partial\phi}{\partial x_3}\right)^4 = \frac{\partial^7\phi}{\partial x_1^3\partial x_3^4}.$$

Remark. *The space $\mathcal{C}^\infty(\mathbb{R}^n)$ consists of all functions $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $D^\alpha\phi \in \mathcal{C}(R^n)$ for each multi–index α. Thus all the partial derivatives of ϕ of all orders exist and are continuous.*

Definition 5.1.4 (Support). *The* **support** *of a function ϕ on a space X is the set*

$$\operatorname{supp}(\phi) := \overline{\{x \in X : \phi(x) \neq 0\}}.$$

Definition 5.1.5 (Space of Test Functions). *The vector space $\mathfrak{D}$ also denoted $\mathcal{C}_c^\infty(\mathbb{R}^n)$, called the* **space of test functions**, *is the space of all functions in $\mathcal{C}^\infty(\mathbb{R}^n)$ with compact support.*

We now show that $\mathfrak{D}$ is nonempty.

Lemma 5.1.6. *For any polynomial p, the function $f : \mathbb{R} \to \mathbb{R}$ defined by*

$$f(x) := \begin{cases} p(1/x)e^{-1/x}, & x > 0, \\ 0, & x \leq 0, \end{cases}$$

is in $\mathcal{C}^\infty(\mathbb{R})$.

Proof. We first show that f is continuous. Note that f is clearly continuous at all nonzero points, so we show continuity at zero. We have

$$\lim_{x\to 0^+} f(x) = \lim_{x\to 0^+} \frac{p(1/x)}{e^{1/x}} = \lim_{t\to+\infty} \frac{p(t)}{e^t}.$$

Through repeated application of L'Hôpital's Rule on the RHS, we see that

$$\lim_{x\to 0^+} f(x) = 0.$$

Thus f is continuous at zero.

Next, differentiation of f gives

$$f'(x) = \begin{cases} Q(1/x)e^{-1/x}, & x > 0, \\ 0, & x < 0, \end{cases}$$

where $Q(1/x) := \frac{1}{x^2}[p(x) - p'(x)]$, by the Chain Rule. By the above argument, $\lim_{x \to 0^+} f'(x) = 0$. We show that $f'(0) = 0$. We have by the Mean Value Theorem that

$$f'(0) = \lim_{h \to 0} \frac{f(h) - f(0)}{h - 0} = \lim_{h \to 0} f'(\xi(h)) = 0$$

for some $\xi(h) \in (0, h)$. Thus it follows

$$f'(x) = \begin{cases} Q(1/x)e^{-1/x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

This is the same form as f, and therefore f' is continuous. This argument can be repeated indefinitely, which completes the proof. $\square$

Lemma 5.1.7. *The function*

$$\rho(x) := \begin{cases} Ce^{\frac{1}{\|x\|^2 - 1}}, & x \in \mathbb{R}^n, \quad \|x\| < 1, \\ 0, & x \in \mathbb{R}^n, \quad \|x\| \geq 1, \end{cases}$$

where $C \in \mathbb{R}$ is chosen so that $\int \rho(x)\ dx = 1$, is an element of $\mathfrak{D}$.

Proof. Take $p(x) \equiv 1$ in the preceding lemma (5.1.6), and note that $\rho(x) = Cf(1 - \|x\|^2)$. Thus $\rho = Fc \circ g$, where $g(x) = 1 - \|x\|^2$, and belongs to $\mathcal{C}^\infty(\mathbb{R}^n)$, since the composition of continuous functions is continuous. By the Chain Rule, $D^\alpha \rho$ may be expressed as a sum of products of ordinary derivatives of f with various partial derivatives of g. Since these are all continuous, $D^\alpha \rho \in \mathcal{C}(\mathbb{R}^n)$ for all multi–indices α.

By the construction, ρ clearly has compact support. Thus $\rho \in \mathfrak{D}$. $\square$

Definition 5.1.8 (Mollifier)**.** *A* **mollifier** *is a function $\phi \in \mathcal{C}^\infty(\mathbb{R}^n)$ such that*

$$\phi \geq 0, \quad \int \phi(x)\ dx = 1, \quad \operatorname{supp}(\phi) \subseteq \{x \in \mathbb{R}^n : \|x\| \leq 1\}.$$

Remark.

(1) The function ρ as defined in Lemma (5.1.7) *is a mollifier.*
(2) If ϕ is a mollifier, then the scaled versions of ϕ, defined by

$$\phi_j(x) = j^n \phi(jx), \quad \{x \in \mathbb{R}^n, j \in \mathbb{N}\},$$

are also mollifiers.

Definition 5.1.9 (Convergence in $\mathfrak{D}$)**.** *A sequence $\{\phi_j\}_{j=1}^\infty$ in $\mathfrak{D}$* **converges** *to zero if the following two conditions are satisfied:*

(1) There exists a compact set $K \subseteq \mathbb{R}^n$ such that $\operatorname{supp}(\phi_j) \subseteq K$ for all $j \in \mathbb{N}$;
(2) For any multi–index α, $D^\alpha \phi_j$ converges uniformly to zero on K.

Moreover, we write $\phi_j \twoheadrightarrow 0$ if $\{\phi_j\}_{j=1}^\infty$ converges to zero in $\mathfrak{D}$. We write $\phi_j \twoheadrightarrow \phi$ if and only if $\phi_j - \phi \twoheadrightarrow 0$.

Recall that uniform convergence to zero on K of the sequence $\{D^\alpha \phi_j\}_{j=1}^\infty$ for any multi–index α means that

$$\sup_{x \in K} |(D^\alpha \phi_j)(x)| \to 0$$

as $j \to \infty$. Since all ϕ_j also vanish outside of K, we have that

$$\sup_{x \in \mathbb{R}^n} |(D^\alpha \phi_j)(x)| \to 0$$

as $j \to \infty$.

Definition 5.1.10 (Continuity on $\mathfrak{D}$)**.** *Let Y be any topological space, and let $F : \mathfrak{D} \to Y$. Then F is continuous if for any sequence $\{\phi_j\}_{j=1}^\infty$ converging to zero in $\mathfrak{D}$, we have that $\{F(\phi_j)\}_{j=1}^\infty$ converges to $F(\phi)$.*

Theorem 5.1.11. *For every multi–index α, D^α is a continuous linear operator of $\mathfrak{D}$ into $\mathfrak{D}$.*

Proof. The linearity of D^α follows immediately from the linearity of differentiation.

For the continuity of D^α, it suffices to show continuity at zero by the linearity of D^α. Thus, suppose that $\{\phi_j\}_{j=1}^\infty$ is a sequence in $\mathfrak{D}$ such that $\phi_j \twoheadrightarrow \phi$. Then there exists a compact set $K \subseteq \mathbb{R}^n$ such that $\mathrm{supp}(\phi_j) \subseteq K$ for all $j \in \mathbb{N}$. by definition, $D^\beta \phi_j(x)$ converges uniformly to zero on K for every multi–index β. Consequently, we have that

$$D^\beta D^\alpha \phi_j$$

converges to zero uniformly on K for every multi–index β, so that $D^\alpha \phi_j \twoheadrightarrow 0$, by definition of convergence in $\mathfrak{D}$. □

Definition 5.1.12 (Distribution)**.** *A* **distribution** *is a continuous linear functional on $\mathfrak{D}$.*

Definition 5.1.13 (Continuity of a Distribution)**.** *Let $T : \mathfrak{D} \to \mathbb{R}$ be a linear functional. Then T is continuous if for any sequence $\{\phi_j\}_{j=1}^\infty$ in $\mathfrak{D}$ such that $\phi_j \twoheadrightarrow 0$, we have that $\{T(\phi_j)\}_{j=1}^\infty$ converges to $0 \in \mathbb{R}$.*

Definition 5.1.14 (Space of Distributions)**.** *We denote the space of all distributions by $\mathfrak{D}'$.*

Example 5.1.15. *A* **Dirac** *distribution δ_ξ is defined by selecting $\xi \in \mathbb{R}^n$ and putting*

$$\delta_\xi(\phi) = \phi(\xi)$$

for all $\phi \in \mathfrak{D}$. This is a distribution, for linearity is satisfied:

$$\begin{aligned} \delta_\xi(\alpha\phi_1 + \beta\phi_2) &= (\alpha\phi_1 + \beta\phi_2)(\xi) = \alpha\phi_1(\xi) + \beta\phi_2(\xi) \\ &= \alpha\delta_\xi(\phi_1) + \beta\delta_\xi(\phi_2). \end{aligned}$$

Secondly, it is continuous because $\phi_j \twoheadrightarrow$ implies immediately that $\phi_j(\xi) \to 0$.

Example 5.1.16. *For $n = 1$, the* **Heaviside** *distribution is defined by*

$$\widetilde{H}(\phi) := \int_0^\infty \phi(x)\ dx,$$

for all $\phi \in \mathfrak{D}$.

Example 5.1.17. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. With f we associate a distribution $\widetilde{f}$ as follows:*

$$\widetilde{f}(\phi) := \int_{\mathbb{R}^n} f(x)\phi(x)\ dx,$$

for all $\phi \in \mathfrak{D}$.

The linearity of $\widetilde{f}$ follows from the linearity of the integral. For continuity, suppose that $\phi_j \twoheadrightarrow 0$. Then there is a compact set K containing all the supports of each ϕ_j, $j \in \mathbb{N}$. Then it follows

$$\lim_{j\to\infty} |\widetilde{f}(\phi_j)| = \lim_{j\to\infty} \left| \int_K f(x)\phi_j(x)\ dx \right| \leq \lim_{j\to\infty} \max_{x\in K} |\phi_j(x)| \int_K |f(x)|\ dx = 0,$$

since $\phi_j \twoheadrightarrow 0$ implies that $\lim_{j\to\infty} \max_{x\in K} |\phi_j(x)| = 0$.

Example 5.1.18. *Fix a multi–index α and define*

$$T(\phi) := \int_{\mathbb{R}^n} D^\alpha \phi(x)\ dx$$

for all $\phi \in \mathfrak{D}$. Then T is a distribution, which follows from Theorem (5.1.11) and Example 3.

Theorem 5.1.19. *If $f \in \mathcal{C}(\mathbb{R}^n)$, then $\widetilde{f} : \mathfrak{D} \to \mathbb{R}$, defined by*

$$\widetilde{f}(\phi) := \int_{\mathbb{R}^n} f(x)\phi(x)$$

for all $\phi \in \mathfrak{D}$, is a distribution. the mapping $T(f) = \widetilde{f}$ is linear and injective from $\mathcal{C}(\mathbb{R}^n)$ into $\mathfrak{D}'$.

Proof. It follows from Example 3 that $\widetilde{f}$ is a distribution.

The linearity of the mapping $T(f) = \widetilde{f}$ follows from the computation

$$\begin{aligned}
(\widetilde{\alpha_1 f_1 + \alpha_2 f_2})(\phi) &= \int_{\mathbb{R}^n} (\alpha_1 f_1 + \alpha_2 f_2)(x)\phi(x)\ dx \\
&= \alpha_1 \int_{\mathbb{R}^n} f_1(x)\phi(x)\ dx + \alpha_2 \int_{\mathbb{R}^n} f_2(x)\phi(x)\ dx \\
&= \alpha_1 \widetilde{f}_1(\phi) + \alpha_2 \widetilde{f}_2(\phi) \\
&= (\alpha_1 \widetilde{f}_1 + \alpha_2 \widetilde{f}_2)(\phi).
\end{aligned}$$

For the injectivity, we recall that it suffices to show that if $f \neq 0$, then $\widetilde{f} \neq 0$. Thus, assume that $f \neq 0$, and let $\xi \in \mathbb{R}^n$ be a point such that $f(\xi) \neq 0$. Select j such that $f(x)$ is of constant sign in the open ball $B(\xi, 1/j)$. Then $\rho_j(x - \xi) = j^n \rho(jx)$, where ρ is defined as in Lemma (5.1.7), is positive in this same ball about ξ and vanishes elsewhere. Hence, $\int_{\mathbb{R}^n} f(x)\rho_j(x - \xi)\ dx \neq 0$. This shows $\widetilde{f}(\phi) \neq 0$, completing the proof. ☐

Note that Example 3 shows that in a certain natural way, each continuous function $f : \mathbb{R}^n \to \mathbb{R}$ "is" a distribution. That is, we can associate a distribution $\widetilde{f}$ with f. In fact, we can extend this notion to some functions that are not continuous.

Definition 5.1.20 (Locally Integrable Function)**.** *A Lebesgue–measurable function $f : \mathbb{R}^n \to \mathbb{R}$ is called* **locally integrable** *if for every compact set $K \subseteq \mathbb{R}^n$, we have $\int_K |f(x)|\ dx < \infty$.*

Remark. *We denote by* $L^1_{loc}(\mathbb{R}^n)$ *the space of equivalence classes of locally integrable functions.*

Theorem 5.1.21. *If f is locally integrable, then the equation $\widetilde{f}(\phi) = \int f\phi \, d\mu$ defines a distribution $\widetilde{f}$ that does not depend on the representative selected from the equivalence class of f. The mapping $T(f) = \widetilde{f}$ is linear and injective from $L^1_{loc}(\mathbb{R}^n)$ into $\mathfrak{D}'$.*

Theorem 5.1.22. *Let μ be any Borel measure on $\mathbb{R}^n$ such that $\mu(K) < \infty$ for each compact set $K \subseteq \mathbb{R}^n$. Then μ induces a distribution T by the equation*

$$T(\phi) := \int_{\mathbb{R}^n} \phi \, d\mu,$$

for all $\phi \in \mathfrak{D}$.

Proof. The linearity of T follows immediately from the linearity of integration.

For the continuity of T, let $\{\phi_j\}_{n=1}^{\infty}$ be a sequence in $\mathfrak{D}$ such that $\phi_j \twoheadrightarrow 0$. Then there exists a compact set $K \subseteq \mathbb{R}^n$ such that $\operatorname{supp}(\phi_j) \subseteq K$ for all $j \in \mathbb{N}$. Consequently, we observe

$$\begin{aligned}\lim_{j\to\infty} |T(\phi_j)| &= \lim_{j\to\infty} \left|\int_K \phi_j \, d\mu\right| \leq \lim_{j\to\infty} \int_K |\phi_j| \, d\mu \leq \lim_{j\to\infty} \max_{x\in K} |\phi_j(x)| \int_K \, d\mu \\ &= \mu(K) \lim_{j\to\infty} \max_{x\in K} |\phi_j(x)| = 0.\end{aligned}$$

This completes the proof. □

Definition 5.1.23 (Regular Distribution)**.** *A distribution of the form $\widetilde{f}(\phi) = \int f\phi \, d\mu$, where $f \in L^1_{loc}(\mathbb{R}^n)$, is called a* **regular distribution***.*

5.2. **Derivatives of Distributions.** We note here that the space $\mathfrak{D}'$ of distributions is very large – it contains images of all continuous functions of $\mathbb{R}^n$ as well as all locally integrable functions. It also contains functionals on $\mathfrak{D}$ that are not readily associated with functions. For instance, the Dirac distribution is a "point evaluation" functional. We now define derivatives of distributions, and show that the new notion of this derivative will coincide with the classical definition when both are defined.

Definition 5.2.1 (Derivative of a Distribution)**.** *If $T \in \mathfrak{D}'$ is a distribution and α any multi–index, then the* **derivative** *of T $\partial^\alpha T$ is the distribution defined by*

$$\partial^\alpha T := (-1)^{|\alpha|} T \circ D^\alpha$$

Lemma 5.2.2. *Let T be a distribution and α any multi–index. Then $\partial^\alpha T$ is a distribution.*

Proof. The linearity of $\partial^\alpha T$ follows from the linearity of T and D^α.

To see that $\partial^\alpha T$ is continuous, let $\{\phi_j\}_{j=1}^{\infty}$ be any sequence in $\mathfrak{D}$ such that $\phi_j \twoheadrightarrow 0$. Then, since T is linear and continuous, we have by continuity of D^α by Theorem (5.1.11) that

$$\begin{aligned}\lim_{j\to\infty} \partial^\alpha T(\phi_j) &= \lim_{j\to\infty} (-1)^{|\alpha|} T(D^\alpha(\phi_j)) = (-1)^{|\alpha|} T\left(\lim_{j\to\infty} D^\alpha(\phi_j)\right) = (-1)^{|\alpha|} T(0) \\ &= 0,\end{aligned}$$

which completes the proof. □

Lemma 5.2.3. *Let $f \in \mathcal{C}^k(\mathbb{R}^n)$. Then $\widetilde{f}$ is a distribution, and for any multi–index α such that $|\alpha| \leq k$,*

$$\partial^\alpha \widetilde{f} = \widetilde{(D^\alpha f)}.$$

Proof. Let $f \in \mathcal{C}^k(\mathbb{R}^n)$ and let α be any multi–index. Without loss of generality, assume that $\alpha_1 \neq 0$. We see that

$$\begin{aligned}\widetilde{(D^\alpha f)}(\phi) &= \int_{\mathbb{R}^n} (D^\alpha f)(x)\phi(x)\ dx \\ &= \int_{\mathbb{R}^n} \left[\left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\phi(x)\ dx.\end{aligned}$$

Write $x' := (x_2, x_3, \ldots, x_n)$. Then

$$\widetilde{(D^\alpha f)}(\phi) = \int_{\mathbb{R}^{n-1}} \int_{-\infty}^{\infty} \left[\left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\phi(x)\ dx_1\ dx'.$$

Integrating by parts once, we obtain

$$\begin{aligned}\widetilde{(D^\alpha f)}(\phi) = \int_{\mathbb{R}^{n-1}} &\left\{ \left[\left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1-1} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\phi(x)\Bigg|_{-\infty}^{\infty} - \right. \\ &\left. \int_{-\infty}^{\infty} \left[\left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1-1} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\left(\frac{\partial \phi}{\partial x_1}\right)(x)\ dx_1 \right\}\ dx' \\ = -\int_{\mathbb{R}^{n-1}} &\int_{-\infty}^{\infty} \left[\left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1-1} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\left(\frac{\partial \phi}{\partial x_1}\right)(x)\ dx_1\ dx',\end{aligned}$$

since ϕ has compact support in $\mathbb{R}^n$ and thus vanishes outside of a closed and bounded set. Repeating this process $\alpha_1 - 1$ times, we find

$$\begin{aligned}\widetilde{(D^\alpha f)}(\phi) &= (-1)^{\alpha_1} \int_{\mathbb{R}^{n-1}} \int_{-\infty}^{\infty} \left[\left(\frac{\partial f}{\partial x_2}\right)^{\alpha_2} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\left(\frac{\partial \phi}{\partial x_1}\right)^{\alpha_1}(x)\ dx_1\ dx' \\ &= (-1)^{\alpha_1} \int_{\mathbb{R}^n} \left[\left(\frac{\partial f}{\partial x_2}\right)^{\alpha_2} \cdots \left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n}\right](x)\left(\frac{\partial \phi}{\partial x_1}\right)^{\alpha_1}(x)\ dx.\end{aligned}$$

Finally, we see that repeating the above process $n-1$ subsequent times gives

$$\begin{aligned}\widetilde{(D^\alpha f)}(\phi) &= (-1)^{|\alpha|} \int_{\mathbb{R}^n} f(x)\left[\left(\frac{\partial \phi}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial \phi}{\partial x_n}\right)^{\alpha_n}\right](x)\ dx \\ &= (-1)^{|\alpha|} \int_{\mathbb{R}^n} f(x) D^\alpha \phi(x) = (-1)^{|\alpha|} \widetilde{f}(D^\alpha \phi) \\ &= \partial^\alpha \widetilde{f}(\phi),\end{aligned}$$

which completes the proof. □

We comment here that it *can* happen that $\partial^\alpha \widetilde{f} \neq \widetilde{(D^\alpha f)}$ for a function f that does not have continuous partial derivatives.

Example 5.2.4. *Let $\widetilde{H}$ be the Heaviside distribution and let δ_0 be the Dirac distribution at zero. Then with $n = 1$ and $\alpha := (1)$, we have $\partial\widetilde{H} = \delta_0$. We observe that for any test function $\phi \in \mathfrak{D}$,*

$$(\partial\widetilde{H})(\phi) = -\widetilde{H}(D\phi) = -\int_0^\infty \phi'(x)\ dx = \phi(0) - \phi(\infty) = \phi(0) = \delta_0(\phi).$$

Example 5.2.5. *Let $n = 1$ and $\alpha := (1)$, so that D is an ordinary derivative. Put*

$$f(x) := \begin{cases} x, & x \geq 0, \\ 0, & , x < 0. \end{cases}$$

Note that f' is not the Heaviside function, since $f'(0)$ is undefined in the classical sense. On the other hand, we see that for any test function, integration by parts gives $\phi \in \mathfrak{D}$,

$$(\partial\widetilde{f})(\phi) = -\widetilde{f}(D\phi) = -\int_{-\infty}^0 f(x)\phi'(x)\ dx - \int_0^\infty f(x)\phi'(x)\ dx = \int_0^\infty \phi(x)\ dx = \widetilde{H}(\phi).$$

That is, in the distributional sense, $f' = H$.

A "distributional derivative" of a function f is a distribution T such that $(\widetilde{f})' = T$. In the general case of an operator D^α, we require $\partial^\alpha\widetilde{f} = T$. If T is a regular distribution, say $T = \widetilde{g}$, then the defining equation is

$$\int g\phi = (-1)^{|\alpha|} \int f D^\alpha \phi,$$

for all $\phi \in \mathfrak{D}$.

Example 5.2.6. *We find the distributional derivative of the function $f(x) = |x|$. We note that it itself is a distribution $\widetilde{g}$, where g is a function such that for all test functions $\phi \in \mathfrak{D}$,*

$$\begin{aligned}
\int_{-\infty}^\infty g(x)\phi(x)\ dx &= -\int_{-\infty}^\infty f(x)\phi'(x)\ dx \\
&= -\int_{-\infty}^0 -x\phi'(x)\ dx - \int_0^\infty x\phi'(x)\ dx \\
&= x\phi(x)|_{-\infty}^0 - \int_{-\infty}^0 \phi(x)\ dx - x\phi(x)|_0^\infty + \int_0^\infty \phi(x)\ dx \\
&= \int_{-\infty}^0 (-1)\phi(x)\ dx + \int_0^\infty (+1)\phi(x)\ dx.
\end{aligned}$$

Hence,

$$g(x) = \begin{cases} -1, & x < 0, \\ 1, & x \geq 0 \end{cases} = 2H(x) - 1,$$

where H denotes the Heaviside function.

We say that $f' = g$ in the distributional sense, or write $\partial\widetilde{f} = \widetilde{g}$. We note that f does not have a classical derivative, because particularly f' is not defined at zero.

Theorem 5.2.7. *The distributional derivative operators ∂^α are linear from $\mathfrak{D}'$ into $\mathfrak{D}'$. Moreover, $\partial^\alpha\partial^\beta = \partial^\beta\partial^\alpha = \partial^{\alpha+\beta}$ for any pair of multi–indices α and β.*

Proof. The linearity of ∂^α has been established in Lemma (5.2.2).

For the commutativity of the operators, we recall that for any function f of two variables, if $\frac{\partial^2 f}{\partial x \partial y}$ and $\frac{\partial^2 f}{\partial y \partial x}$ exist and are continuous, then they are equal. More generally, for any function $f \in \mathcal{C}^k(\mathbb{R}^n)$, up to k partial derivatives of f may be interchanged. Therefore, for any $\phi \in \mathfrak{D}$, we have $D^\alpha D^\beta \phi = D^\beta D^\alpha \phi$. Consequently, for any arbitrary distribution $T \in \mathfrak{D}'$ and test function $\phi \in \mathfrak{D}$, we have

$$\begin{aligned}
\partial^\alpha \partial^\beta T &= \partial^\alpha(\partial^\beta T) = (-1)^{|\alpha|}(\partial^\beta T) \circ D^\alpha = (-1)^{|\alpha|}(-1)^{|\beta|}(T \circ D^\beta) \circ D^\alpha \\
&= (-1)^{|\alpha|+|\beta|} T \circ D^\beta \circ D^\alpha = (-1)^{|\beta|}(-1)^{|\alpha|} T \circ D^\alpha \circ D^\beta \\
&= (-1)^{|\alpha|}(\partial^\alpha T) \circ D^\beta = \partial^\beta \partial^\alpha T,
\end{aligned}$$

which completes the proof. □

Theorem 5.2.8. *For $n = 1$, that is, for functions of one variable, every distribution is the derivative of another distribution.*

Proof. Let $\widetilde{1}$ be the regular distribution induced by the constant 1 : that is,

$$\widetilde{1}(\phi) := \int_{-\infty}^{\infty} \phi(x)\ dx$$

for all $\phi \in \mathfrak{D}$. Note that $\ker(\widetilde{1})$ is a closed hyperplane in $\mathfrak{D}$. Choose a test function $\psi \in \mathfrak{D}$ such that

$$\widetilde{1}(\psi) = \int_{-\infty}^{\infty} \psi(x)\ dx = 1.$$

Define

$$A\phi := \phi - \widetilde{1}\phi\psi$$

for all $\phi \in \mathfrak{D}$, and

$$(B\phi)(x) := \int_{-\infty}^{x} \phi(y)\ dy$$

for all $\phi \in \ker(\widetilde{1})$. We see that if $\phi \in \ker(\widetilde{1})$, then $B\phi \in \mathfrak{D}$, since ϕ has compact support.

Now let $T \in \mathfrak{D}'$ be any distribution, and define $S := -T \circ B \circ A$. We show that S is a distribution and that $\partial S = T$. First note that for all test functions $\phi \in \mathfrak{D}$, we have

$$\widetilde{1}(A\phi) = \widetilde{1}(\phi - \widetilde{1}(\phi)\psi) = \widetilde{1}(\phi) - \widetilde{1}(\phi)\widetilde{1}\psi = \widetilde{1}(\phi)(1-1) = 0,$$

so that $A\phi \in \ker(\widetilde{1})$. Thus, $BA\phi \in \mathfrak{D}$. Since $B \circ A$ is continuous from $\mathfrak{D}$ into $\mathfrak{D}$, and because T is a distribution, it follows that S is a distribution. Finally, we compute

$$\begin{aligned}
(\partial S)(\phi)(x) &= -S(\phi')(x) = T(BA\phi')(x) = TB(A\phi)(x) \\
&= TB\left(\phi'(x) - \widetilde{1}(\phi')\psi(x)\right) = TB\left(\phi'(x) - \left\{\int_{-\infty}^{\infty} \phi'(x)\ dx\right\}\psi(x)\right) \\
&= TB(\phi')(x) = T((B\phi')(x)) = T\left(\int_{-\infty}^{x} \phi'(y)\ dy\right) \\
&= T(\phi)(x).
\end{aligned}$$

This completes the proof. □

Theorem 5.2.9. *Let $n = 1$, and let T be a distribution for which $\partial T = 0$. Then $T = \widetilde{c}$ for some $c \in \mathbb{R}$.*

Proof. We adopt the same notation as the proof of Theorem (5.2.8). Recall from the Fundamental Theorem of Calculus that

$$\phi(x) = \frac{d}{dx} \int_{-\infty}^{x} \phi(y)\ dy,$$

which implies $\phi \equiv (DB)(\phi)$. Note that this equation holds for all $\phi \in \ker(\widetilde{1})$. Moreover, since $A\phi \in \ker(\widetilde{1})$ for all $\phi \in \mathfrak{D}$, we have $A\phi = (DBA)(\phi)$ for all $\phi \in \mathfrak{D}$. Consequently, if $\partial T = 0$, then for all test functions $\phi \in \mathfrak{D}$, we have

$$\begin{aligned} T(\phi) &= T(A\phi + \widetilde{1}(\phi)\psi) = T(A\phi) + \widetilde{1}(\phi)T(\psi) \\ &= T((DBA)(\phi)) + \widetilde{1}(\phi)T(\psi) \\ &= -(\partial T)(BA(\phi)) + T(\psi)\widetilde{1}(\phi) \\ &= T(\psi)\widetilde{1}(\phi) = T(\psi) \int_{-\infty}^{\infty} \phi(x)\ dx \\ &= \int_{-\infty}^{\infty} T(\psi)\phi(x)\ dx. \end{aligned}$$

Thus $T = \widetilde{c}$, with $c = T(\psi)$. □

Theorem 5.2.10. *If $T \in \mathfrak{D}'$ is a distribution and $K \subseteq \mathbb{R}^n$ is a compact set in $\mathbb{R}^n$, then there exists $f \in \mathcal{C}(\mathbb{R}^n)$ and a multi–index α such that for all $\phi \in \mathfrak{D}$ with* $\operatorname{supp}(\phi) \subseteq K$,

$$T(\phi) = (\partial^\alpha \widetilde{f})(\phi).$$

5.3. Convergence of Distributions.

Definition 5.3.1 (Convergence in $\mathfrak{D}'$)**.** *Let $\{T_j\}_{j=1}^{\infty} \subseteq \mathfrak{D}'$ be a sequence of distributions. We say that $\{T_j\}_{j=1}^{\infty}$ converges to zero if for every $\phi \in \mathfrak{D}$, we have that $\{T_j(\phi)\}_{j=1}^{\infty}$ converges to zero. We write*

$$T_j \to 0 \quad \Longleftrightarrow \quad T_j(\phi) \to 0 \quad \forall \phi \in \mathfrak{D}.$$

Moreover, we say that $\{T_j\}_{j=1}^{\infty}$ converges to a distribution $T \in \mathfrak{D}'$ if for all $\phi \in \mathfrak{D}$, $\{(T_j - T)(\phi)\}_{j=1}^{\infty}$ converges to zero.

Notice that this definition coincides with weak* convergence of a sequence of linear functionals. Topological notions in $\mathfrak{D}'$, such as continuity, will be based on this notion of convergence.

Theorem 5.3.2. *For every multi–index α, ∂^α is a continuous linear map of $\mathfrak{D}'$ into $\mathfrak{D}$.*

Proof. Let $\{T_j\}_{j=1}^{\infty}$ be a sequence in $\mathfrak{D}'$ such that $T_j \to 0$. Let $\phi \in \mathfrak{D}$ be any test function. Recall that, for any $j \in \mathbb{N}$,

$$\partial^\alpha T_j(\phi) = (-1)^{|\alpha|} T_j(D^\alpha \phi)$$

Since $D^\alpha \phi$ is a test function, it follows by the assumption that

$$\partial^\alpha T_j(\phi) = (-1)^{|\alpha|} T_j(D^\alpha \phi) \to 0$$

as $j \to \infty$. This completes the proof. $\square$

Theorem 5.3.3. *If a sequence of distributions $\{T_j\}_{j=1}^\infty$ has the property that $\{T_j(\phi)\}_{j=1}^\infty$ is convergent for each test function ϕ, then the equation $T(\phi) := \lim_{j\to\infty} T_j(\phi)$ defines a distribution T, and $T_j \to T$.*

Proof. Let $\{T_j\}_{j=1}^\infty$ be a sequence in $\mathfrak{D}'$ such that $\lim_{j\to\infty} T_j(\phi)$ exists in $\mathbb{R}$ for all $\phi \in \mathfrak{D}$. Define a mapping T by

$$T(\phi) := \lim_{j\to\infty} T_j(\phi).$$

Note that T is clearly well–defined.

To see that T is linear, we have for all $\alpha, \beta \in \mathbb{R}$ and $\phi, \psi \in \mathfrak{D}$ that

$$\begin{aligned} T(\alpha\phi + \beta\psi) &= \lim_{j\to\infty} T_j(\alpha\phi + \beta\psi) = \lim_{j\to\infty} [\alpha T_j(\phi) + \beta T_j(\psi)] \\ &= \alpha \lim_{j\to\infty} T_j(\phi) + \beta \lim_{j\to\infty} T_j(\psi) \\ &= \alpha T(\phi) + \beta T(\psi). \end{aligned}$$

For the continuity of T, see Rudin's Functional Analysis. $\square$

Corollary 5.3.4. *A series of distributions $\sum_{j=1}^\infty T_j$ converges to a distribution $T \in \mathfrak{D}'$ if and only if for each test function $\phi \in \mathfrak{D}$, the series $\sum_{j=1}^\infty T_j(\phi)$ converges in $\mathbb{R}$.*

Proof. Define a new sequence of distributions $\{S_N\}_{N=1}^\infty$ by

$$S_N(\phi) := \sum_{j=1}^N T_j(\phi).$$

By the assumption, $\{S_N(\phi)\}_{N=1}^\infty$ is convergent in $\mathbb{R}$ for each test function $\phi \in \mathfrak{D}$, so that by Theorem (5.3.3) we have that $\{S_N\}_{N=1}^\infty$ converges to a distribution $T \in \mathfrak{D}'$. $\square$

Corollary 5.3.5. *Let $\{T_j\}_{j=1}^\infty$ be a sequence in $\mathfrak{D}'$ such that $\sum_{j=1}^\infty T_j$ converges. Then for any multi–index α, $\partial^\alpha \sum_{j=1}^\infty T_j = \sum_{j=1}^\infty \partial^\alpha T_j$.*

Proof. By Theorem (5.3.2), ∂^α is a continuous linear operator. Hence, for all test functions $\phi \in \mathfrak{D}$,

$$\begin{aligned} \partial^\alpha \left(\sum_{j=1}^\infty T_j \right) &= \partial^\alpha \left(\lim_{N\to\infty} \sum_{j=1}^N T_j \right) = \lim_{N\to\infty} \left(\partial^\alpha \sum_{j=1}^N T_j \right) \\ &= \lim_{N\to\infty} \sum_{j=1}^N \partial^\alpha T_j = \sum_{j=1}^\infty \partial^\alpha T_j. \end{aligned}$$

$\square$

Note the contrast with the above results and the results for the classical derivative. Recall that a pointwise convergent sequence of continuous functions need not have a continuous limit. For example, consider the sequence $\{f_n\}_{n=1}^{\infty} \subseteq \mathcal{C}[0,1]$ defined by

$$f_n(t) := \begin{cases} 1-n, & 0 \le t < \frac{1}{n}, \\ 0, & \frac{1}{n} \le t \le 1. \end{cases}$$

It may be shown that this sequence converges pointwise to the function

$$f(t) := \begin{cases} 0, & t = 0, \\ 1, & 0 < t \le 1. \end{cases}$$

Similarly, even a *uniformly* convergent series of continuously differentiable functions can fail to satisfy the equation

$$\frac{d}{dx}\sum_{n=1}^{\infty} F_n = \sum_{n=1}^{\infty} \frac{d}{dx} f_n.$$

A famous example is provided by the Weierstrass function

$$f(x) := \sum_{n=1}^{\infty} 2^{-n}\cos(3^n x).$$

This function is continuous but differentiable nowhere.

Example 5.3.6. *Define the sequence $\{f_n(x)\}_{n=1}^{\infty}$ by $f_n(x) := \cos(nx)$ for each $n \in \mathbb{N}$. Note that this sequence of functions does not converge even pointwise. On the other hand, the sequence of distributions $\{\widetilde{f}_n\}_{n=1}^{\infty}$ converges to zero. We observe that, for any test function $\phi \in \mathfrak{D}(\mathbb{R})$,*

$$\begin{aligned} \lim_{n\to\infty} |\widetilde{f}_n(\phi)| &= \lim_{n\to\infty} \left| \int_{-\infty}^{\infty} f_n \phi \right| = \lim_{n\to\infty} \left| \int_{-\infty}^{\infty} \cos(nx)\phi(x)\ dx \right| \\ &= \lim_{n\to\infty} \left\{ \left| \frac{1}{n}\sin(nx)\phi(x) \right|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{n}\sin(nx)\phi'(x)\ dx \right| \right\} \\ &= \lim_{n\to\infty} \left| \frac{1}{n}\int_{-\infty}^{\infty} \sin(nx)\phi'(x)\ dx \right|. \end{aligned}$$

Since $\phi'(x)$ is a test function, it has compact support. Thus there exists a closed interval $[a,b]$ such that $\operatorname{supp}(\phi') \subseteq [a,b]$. *Hence,*

$$\begin{aligned} \lim_{n\to\infty} |\widetilde{f}_n(\phi)| = \lim_{n\to\infty} \left| \frac{1}{n}\int_a^b \sin(nx)\phi'(x)\ dx \right| &\le \lim_{n\to\infty} \frac{1}{n}\int_a^b |\sin(nx)\phi'(x)|\ dx \\ &\le \lim_{n\to\infty} \frac{1}{n}\left\{(b-a)\max_{x\in[a,b]}|\phi'(x)|\right\} \\ &= 0. \end{aligned}$$

Theorem 5.3.7. *Let $\{f_n\}_{n=1}^{\infty}$ be a sequence in $L^1_{loc}(\mathbb{R}^n)$, and suppose that $\{f_n\}_{n=1}^{\infty}$ converges pointwise to a function $f \in L^1_{loc}(\mathbb{R}^n)$ almost everywhere. If there exists an element $g \in L^1_{loc}(\mathbb{R}^n)$ such that $|f_j| \le g$ for all $j \in \mathbb{N}$, then $\widetilde{f}_j \to \widetilde{f}$ in $\mathfrak{D}'$.*

Proof. Let $\phi \in \mathfrak{D}(\mathbb{R}^n)$ be any test function, and let $K := \operatorname{supp}(\phi)$. Note that, for all $j \in \mathbb{N}$, we have $f_j\phi \in L^1(K)$. Furthermore, we have that

$$|f_j\phi| = |f_j||\phi| \leq g|\phi|$$

for each $j \in \mathbb{N}$ and $(f_j \circ \phi)(x) \to (f \circ \phi)(x)$ pointwise almost everywhere. Hence, by the Dominated Convergence Theorem, it follows

$$\begin{aligned}\lim_{j\to\infty} \widetilde{f}_j(\phi) &= \lim_{j\to\infty} \int f_j\phi \; d\mu = \int \lim_{j\to\infty} f_j\phi \; d\mu = \int f\phi \; d\mu \\ &= \widetilde{f}(\phi).\end{aligned}$$

This is for all $\phi \in \mathfrak{D}(\mathbb{R}^n)$, so that $\lim_{j\to\infty} \widetilde{f}_j = \widetilde{f}$. □

Theorem 5.3.8. *Let $\{f_j\}_{j=1}^\infty$ be a sequence of nonnegative functions in $L^1_{loc}(\mathbb{R}^n)$ such that $\int f_j \; d\mu = 1$ for each $j \in \mathbb{N}$ and such that*

$$\lim_{j\to\infty} \int_{[|x|\geq r]} f_j \; d\mu = 0$$

for all $r > 0$. Then $\widetilde{f}_j \to \delta$, the Dirac distribution.

Proof. Let $\phi \in \mathfrak{D}(\mathbb{R}^n)$ be any test function, and put $\psi := \phi - \phi(0)$. Fix $\epsilon > 0$, and select $r > 0$ so that $|\psi(x)| < \epsilon$ whenever $|x| < r$. Since $\int f_j \; d\mu = 1$ for all $j \in \mathbb{N}$, we see that

$$\begin{aligned}\lim_{j\to\infty} |\widetilde{f}_j(\phi) - \delta| &= \lim_{j\to\infty} = \left\{\left|\int_{\mathbb{R}^n} f_j\phi \; d\mu - \phi(0)\right|\right\} = \lim_{j\to\infty} \left\{\left|\int_{\mathbb{R}^n} f_j\phi \; d\mu - \phi(0)\int_{\mathbb{R}^n} f_j \; d\mu\right|\right\} \\ &= \lim_{j\to\infty} \left|\int_{\mathbb{R}^n} f_j(\phi - \phi(0)) \; d\mu\right| = \lim_{j\to\infty} \left|\int_{\mathbb{R}^n} f_j\psi \; d\mu\right| \\ &\leq \lim_{j\to\infty} \int_{\mathbb{R}^n} |f\psi| \; d\mu = \lim_{j\to\infty} \left\{\int_{|x|<r} |f_j\psi| \; d\mu + \int_{|x|\geq r} |f_j\psi| \; d\mu\right\} \\ &< \lim_{j\to\infty} \left\{\epsilon\int_{\mathbb{R}^n} f_j \; d\mu + \max_{x\in\operatorname{supp}(\psi)} |\psi(x)| \int_{|x|\geq r} f_j \; d\mu\right\} \\ &= \epsilon \lim_{j\to\infty} \int_{\mathbb{R}^n} f_j \; d\mu + \max_{x\in\operatorname{supp}(\psi)} |\psi(x)| \lim_{j\to\infty} \int_{|x|\geq r} f_j \; d\mu \\ &= \epsilon.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary, it follows that $\lim_{j\to\infty} \widetilde{f}_j(\phi) = \delta(\phi)$ for all $\phi \in \mathfrak{D}(\mathbb{R}^n)$. Hence, $\widetilde{f}_j \to \delta$. □

5.4. **Convolutions.**

Definition 5.4.1 (Convolution)**.** *The* **convolution** *of two functions f and ϕ on $\mathbb{R}^n$ is a function $f * \phi$ defined by*

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

Note that the above integral exists and is finite if $\phi \in \mathfrak{D}$ and $f \in L^1_{loc}(\mathbb{R}^n)$, because for each $x \in \mathbb{R}^n$, the integration takes place over a compact subset of $\mathbb{R}^n$. With a change of variable in the integral, $z := x - y$, we see that

$$(f * \phi)(x) = \int_{\mathbb{R}^n} f(x-z)\phi(z)\, dz = (\phi * f)(x).$$

In taking the convolution of two functions, we expect that some favorable properties of one function will be inherited by the convolution function. For example, suppose that f is integrable, and let $\phi \in \mathfrak{D}$. In the case $n = 1$, differentiating $f * \phi$ with respect to x gives

$$(f * \phi)'(x) = \int_{-\infty}^{\infty} f(y)\phi'(x-y)\, dy.$$

Thus the differentiability of ϕ is inherited by the convolution $f * \phi$. This property also holds with higher derivatives and with functions of several variables.

It follows from the above that if ϕ is a polynomial of degree at most k, then so is $f * \phi$, for the $(k+1)-$st derivative of $f * \phi$ will be zero. Similarly, if ϕ is a periodic function, then so is $f * \phi$.

We will see that convolutions are useful in approximating functions by smooth functions. Here mollifiers play a role. Recall that a mollifier is a function $\phi \in \mathfrak{D}$ such that $\phi \geq 0$, $\int \phi = 1$, and $\phi(x) = 0$ when $|x| \geq 1$. Let ϕ be a mollifier. Define $\phi_j(x) := j^n \phi(jx)$. Since $\int \phi_j(x) = 1$,

$$\begin{aligned} f(x) - (f * \phi_j)(x) &= f(x) - \int f(x-z)\phi_j(z)\, dz \\ &= \int f(x)\phi_j(z)\, dz - \int f(x-z)\phi_j(z)\, dz \\ &= \int [f(x) - f(x-z)]\phi_j(z)\, dz. \end{aligned}$$

Since ϕ vanishes outside the unit ball in $\mathbb{R}^n$, ϕ_j vanishes outside the ball of radius $1/j$. Hence in the equation above the only values of z that have any effect are those for which $|z| < 1/j$. If f is uniformly continuous, the calculation shows that $f * \phi_j(x)$ is close to $f(x)$, and we have therefore approximated f by the smooth function $f * \phi$.

Define linear operators E_x and B by

$$\begin{aligned} (E_x\phi)(y) &:= \phi(y - x), \\ (B\phi)(y) &:= \phi(-y). \end{aligned}$$

Thus we have

$$(f * \phi)(x) = \widetilde{f}(E_x B\phi).$$

For $f \in L^1_{loc}(\mathbb{R}^n)$ and $\phi \in \mathfrak{D}$ we have

$$\begin{aligned} \widetilde{E_x f}(\phi) &= \int E_x f * \phi = \int f(y-x)\phi(y)\, dy = \int f(z)\phi(z+x)\, dz \\ &= \widetilde{f}(E_{-x}\phi). \end{aligned}$$

Definition 5.4.2. *If T is a distribution, we define $E_x T := T E_{-x}$. If $\phi \in \mathfrak{D}$, then the convolution $T * \phi$ is defined by $(T * \phi)(x) := T(E_x B\phi)$.*

Lemma 5.4.3. *For $T \in \mathfrak{D}'$ and $\phi \in \mathfrak{D}$,*

$$E_x(T * \phi) = (E_x T) * \phi = T * E_x\phi.$$

Proof. By straightforward calculation,

$$[E_x(T * \phi)(y)] = (T * \phi)(y - x) = T(E_{y-x}B\phi),$$
$$[(E_x)T * \phi](y) = (E_xT)(E_yB\phi) = T(E_{-x}E_yB\phi) = T(E_{y-x}B\phi),$$
$$[T * E_x\phi](y) = T(E_yBE_x\phi) = T(E_yE_{-x}B\phi) = T(E_{y-x}B\phi).$$

□

Lemma 5.4.4. *If T is a distribution and if $\phi_j \twoheadrightarrow \phi$ in $\mathfrak{D}$, then $T * \phi_j \to T * \phi$ pointwise.*

Proof. By linearity, it suffices to consider the case $\phi = 0$. If $\phi_j \twoheadrightarrow 0$ in $\mathfrak{D}$, then, for all $x \in \mathbb{R}^n$,

$$(T * \phi_j)(x) = T(E_xB\phi_j) \to 0$$

by continuity of B, E_x, and T. □

Lemma 5.4.5. *Let $\{x_j\}_{j=1}^{\infty}$ be a sequence of points in $\mathbb{R}^n$ converging to x. For each $\phi \in \mathfrak{D}$,*

$$E_{x_j}\phi \twoheadrightarrow E_x\phi$$

in $\mathfrak{D}$.

Proof. Choose $\phi \in \mathfrak{D}$, let $K_1 := \{x_j\}_{j=1}^{\infty} \cup \{x\}$, and let $K_2 := \mathrm{supp}\phi$. Then the supports of $E_{x_j}\phi$ are contained in the compact set

$$K_1 + K_2 := \{u + v : u \in K_1, v \in K_2\}.$$

Next, we observe that $(E_{x_j}\phi)(y) \to (E_x\phi)(y)$ uniformly for $y \in K_1 + K_2$. Indeed, by the uniform continuity of ϕ over a compact set, for any given $\epsilon > 0$, there exists $\delta > 0$ such that $|\phi(u) - \phi(v)| < \epsilon$ whenever $|u - v| < \delta$. Hence, if $|x_j - x| < \delta$, then $|\phi(y - x_j) - \phi(y - x)| < \epsilon$. Thus it follows that

$$(D^\alpha E_{x_j}\phi)(y) \to (D^\alpha E_x\phi)(y)$$

uniformly for $y \in K_1 + K_2$, because $D^\alpha E_{x_j}\phi = E_{x_j}D^\alpha\phi$, and the above argument may be applied to $D^\alpha\phi$. □

6. Additional Topics

6.1. **Compact Operators and the Fredholm Theory.** This section is focused on "perturbations of the identity," that is, operators $I + A$, where I is the identity operator and A is a compact operator. Recall that in §2.3, we found that operators with finite–dimensional range are compact, and that the set of compact operators in $\mathcal{L}(X, Y)$ is closed if X and Y are Banach spaces. Thus, the limit of operators, each having finite–dimensional range, is necessarily a compact operator. In many Banach spaces, every compact operator is such a limit. This fact can be exploited in many practical problems involving compact operators — we may approximate the operator by a simpler one having finite–dimensional range.

In this section, we consider a related class of operators, namely, those of the form $I + A$, where A is compact, and find that such operators have favorable properties also. Intuitively, we expect such operators to be well–behaved, because they are close to the identity operator. For instance, we will prove the Fredholm Alternative, which asserts that for such operators the property of injectivity is equivalent to surjectivity. This is a familiar theorem in the context of linear operators from $\mathbb{R}^n$ to $\mathbb{R}^n$: recall that for an $n \times n$ matrix, the properties of having a zero–dimensional kernel and an $n-$dimensional range are equivalent.

Lemma 6.1.1. *Let A be a compact operator on a normed linear space. If $I + A$ is surjective, then it is injective.*

Proof. Put $B := I + A$ and $X_n := \ker(B^n)$. By contradiction, suppose that B is surjective but not injective. By linearity, note that

$$\{0\} \subseteq X_1 \subseteq X_2 \subseteq \cdots .$$

We now show that the above inclusions are proper. By the assumption that B is not injective, there exists a nonzero element $y_1 \in X_1$. Since B is surjective, there exist points $y_2, y_3, \ldots$ such that $By_{n+1} = y_n$ for $n = 2, 3, \ldots$. Note that

$$B^n y_n = B^{n-1} B y_n = B^{n-1} y_{n-1} = \cdots = B^2 y_2 = B y_1 = 0,$$

so that $y_n \in X_n$ for each $n \in \mathbb{N}$. Moreover,

$$B^{n-1} y_n = B^{n-2} B y_n = B^{n-2} y_{n-1} = \cdots = B^2 y_3 = B y_2 = y_1 \neq 0,$$

and thus $y_n \notin X_{n-1}$ for each $n \in \mathbb{N}$. This proves $y_n \in X_n \setminus X_{n-1}$ for all $n \in \mathbb{N}$, and thus the above inclusions are proper.

Next, by the Riesz Lemma, there exist points $x_n \in X_n$ such that $\|x_n\| = 1$ and $\operatorname{dist}(x_n, X_{n-1}) \geq 1/2$. Recall that $B^m x_m = 0$ because $x_m \in X_m = \ker(B^m)$. Moreover, if $m > n$, then $B^{m-1} x_n = 0$ because $x_n \in X_n \subset X_{m-1}$. Finally, $B^m x_n = 0$ because $x_n \in X_n \subset X_m$. Thus

$$B^{m-1}(Bx_m + x_n - Bx_n) = B^m x_m + B^{m-1} x_n - B^m x_n = 0.$$

Thus $Bx_m + x_n - Bx_n \in X_{m-1}$, and for all $m > n$, we can write

$$\begin{aligned} \|Ax_n - Ax_m\| &= \|(B - I)x_n - (B - I)x_m\| = \|Bx_n - x_n - Bx_m + x_m\| \\ &= \|x_m - (Bx_m + x_n - Bx_n)\| \geq \operatorname{dist}(x_m, X_{m-1}) \geq 1/2. \end{aligned}$$

This shows that the sequence $\{Ax_n\}_{n=1}^{\infty}$ can have no Cauchy subsequence, a contradiction to the compactness of A. □

Lemma 6.1.2. *Let A be a compact operator on a Banach space. Then the range of $I + A$ is closed.*

Proof. Put $B := I + A$. Let $\{y_n\}_{n=1}^{\infty}$ be a convergent sequence in the range of B, and write $y := \lim_{n\to\infty} y_n$. We want to show that $y \in \mathcal{R}(B)$. Since this is obvious if $y = 0$, we assume that $y \neq 0$. Denote $K := \ker(B)$, and let $y_n = Bx_n$ for points x_n.

First suppose that $\{x_n\}_{n=1}^{\infty}$ contains a bounded subsequence. Then, since A is compact, $\{Ax_n\}_{n=1}^{\infty}$ contains a convergent subsequence, say $\lim_{i\to\infty} Ax_{n_i} = u$. Since $Ax_{n_i} + x_{n_i} = Bx_{n_i} = y_{n_i}$, we infer that

$$\lim_{i\to\infty} x_{n_i} = \lim_{i\to\infty} \{y_{n_i} - Ax_{n_i}\} = y - u.$$

Thus, by continuity, $y = \lim_{i\to\infty} Bx_{n_i} = B(\lim_{i\to\infty} x_{n_i}) = B(y-u)$, and $y \in \mathcal{R}(B)$. This completes the proof for this case.

Now assume that $\{x_n\}_{n=1}^{\infty}$ contains no bounded subsequence. Then $\lim_{n\to\infty} \|x_n\| = \infty$. Since $y \neq 0$, we may remove finitely many terms from the sequence $\{x_n\}_{n=1}^{\infty}$ and assume that $x_n \notin K$ for all $n \in \mathbb{N}$. Using the Riesz Lemma, construct vectors $v_n := k_n + \alpha_n x_n$ so that $\|v_n\| = 1, k_n \in K$, and $\operatorname{dist}(v_n, K) \geq 1/2$. Note that

$$Bv_n = \alpha Bx_n = \alpha_n y_n.$$

Since $\|\alpha_n y_n\| = \|Bv_n\| \leq \|B\|$ and $\lim_{n\to\infty} y_n = y \neq 0$, the sequence $\{\alpha_n\}_{n=1}^{\infty}$ is bounded. Since $\{v_n\}_{n=1}^{\infty}$ is bounded, $\{Av_n\}_{n=1}^{\infty}$ contains a convergent subsequence. By the Bolzano–Weierstrass Theorem, $\{\alpha_n\}_{n=1}^{\infty}$ also has a convergent subsequence, and, reindexing if necessary, we may assume that

$$\lim_{i\to\infty} Av_{n_i} = z, \qquad \lim_{i\to\infty} \alpha_{n_i} = \alpha.$$

Thus we conclude that $Bv_{n_i} = (I + A)v_{n_i} = \alpha_{n_i} y_{n_i}$ and

$$\lim_{i\to\infty} v_{n_i} = \lim_{i\to\infty} \{\alpha_{n_i} y_{n_i} - Av_{n_i}\} = \alpha y - z.$$

If α were 0, we would have $\lim_{i\to\infty} v_{n_i} = -z$ and thus

$$B(-z) = \lim_{i\to\infty} Bv_{n_i} = \lim_{i\to\infty} (v_{n_i} + Av_{n_i}) = -z + z = 0.$$

This would imply $z \in K$. On the other hand, this would imply

$$1/2 \leq \operatorname{dist}(v_{n_i}, K) \leq \|v_{n_i} + z\| = 0.$$

Hence $\alpha \neq 0$. Since $\lim_{i\to\infty} Bv_{n_i} = \alpha y$, we have

$$\lim_{i\to\infty} B(v_{n_i}/\alpha) = y.$$

Consequently, $B(y - z/\alpha) = y$, and we have $y \in \mathcal{R}(B)$. ☐

Lemma 6.1.3. *Let A be a compact operator on a Banach space. If $I + A$ is injective, then it is surjective.*

Proof. Let $B := I + A$ and let $X_n := \mathcal{R}(B^n)$. By the binomial theorem,

$$B^n = (I + A)^n = \sum_{k=0}^{n} \binom{n}{k} A^k = I + \sum_{k=1}^{n} \binom{n}{k} A^k.$$

Since each A^k is compact for $k \geq 1$, B^n is the identity plus a compact operator. Thus X_n is closed by Lemma 2.

Next let $x \in X_n$ for some $n \in \mathbb{N}$. Then there exists u such that $x = B^n u$, and thus

$$x = B^n u = B^{n-1} Bu \in X_{n-1}.$$

Thus

$$X := X_0 \supseteq X_1 \supseteq X_2 \supseteq \cdots .$$

We now show that $X_0 = X_1$.

By contradiction, suppose that all of the above inclusions are proper. By the Riesz Lemma, there exists $x_n \in X_n$ such that $\|x_n\| = 1$ and $\operatorname{dist}(x_n, X_{n+1}) \geq 1/2$. Then, for $n < m$, we have

$$\begin{aligned}\|Ax_m - Ax_n\| = \|(B-I)x_m - (B-I)x_n\| &= \|x_n - (x_m + Bx_n - Bx_m)\| \\ &\geq \operatorname{dist}(x_n, X_{n+1}) \geq 1/2,\end{aligned}$$

because $x_m \in X_m \subseteq X_{n+1}$, $Bx_m \in X_{m+1} \subset X_{n+1}$, and $Bx_n \in X_{n+1}$. This shows that $\{Ax_n\}_{n=1}^{\infty}$ can contain no Cauchy subsequence, a contradiction to the compactness of A.

Thus not all of the above inclusions are proper, and for some $n \in \mathbb{N}$, $X_n = X_{n+1}$. We define n to be the first integer having this property. We show that $n = 0$.

By contradiction, suppose that $n > 0$. Let $x \in X_{n-1}$ be arbitrary. Then $x = B^{n-1}y$ for some y, and

$$Bx = B^n y \in X_n = X_{n+1}.$$

It follows that $Bx = B^{n+1}z$ for some z. Since B is injective by the hypothesis, $x = B^n z \in X_n$. Since x was an arbitrary point in X_{n-1}, this shows that $X_{n-1} \subseteq X_n$. But the inclusion $X_n \subseteq X_{n-1}$ also holds. Hence $X_{n-1} = X_n$, a contradiction to the choice of n. Thus $n = 0$, and we have $X = \mathcal{R}(B)$. □

Theorem 6.1.4. *Let A be a compact linear operator on a Banach space. Then the operator $I + A$ is injective if and only if it is surjective.*

Proof. See Lemmas 1 and 3. □

The traditional formulation of the above theorem states that one and only one of these alternatives holds:

(1) $I + A$ is surjective;
(2) $I + A$ is not injective.