VECTOR SPACES

MARK H. KIM

Abstract. This set of notes concerns the basic theory of vector spaces. Advanced topics such as module theory ans spectral theory are not covered.

Contents

1. Fields 2
2. Vector Spaces 7
3. Bases, Products, Sums, and Free Vector Spaces 8
4. Matrices 18
5. Quotient Spaces 21
6. Duality 25
7. Tensor Products; Change of Base Field 30
8. Normed Linear Spaces 37
9. Bounded Linear Operators 44
10. Completeness: Applications of the Baire Category Theorem 47
10.1. Nowhere Dense Sets and the Baire Category Theorem 47
10.2. The Open Mapping Theorem 48
10.3. Closed Graph Theorem 50
10.4. Uniform Boundedness Principle 52
11. Duality: The Hahn-Banach theorems 54
12. Convexity, Part 1: the Geometric Hahn-Banach Theorems 60
13. Convexity, Part 2: the Krein-Milman Theorem 64
14. Weak and Weak-$*$ Topologies 67
14.1. The Weak Topology Paradigm and the Product Topology 67
14.2. Weak-$*$ Topology 69
14.3. Weak Topology 72
14.4. Concluding Remarks 73
15. Locally Convex Spaces 75
16. Hilbert Spaces 80
16.1. Inner Products 80
16.2. Orthogonality 83
16.3. Separability 90
16.4. Algebra of Operators 92
17. Appendix: Categories 96
17.1. Basic Definitions 96
17.2. Universality 99
17.3. Natural Transformations 101
18. Appendix: Sets 102
18.1. Order theory 103

18.2. Cardinality 106
19. Appendix: Analytic Miscellanies 112
19.1. Quotient Spaces of a Normed Linear Space 112
19.2. Annihilators 114
19.3. Moore–Smith Theory of Nets 116
References 119

1. Fields

We begin by considering generalizations of the real and complex number systems.

Definition 1.1. A *field* is a quintuplet $(F, +, \cdot, 0_F, 1_F)$ of a set F equipped with two binary operations $+ : F \times F \to F$ and $\cdot : F \times F \to F$ and identity elements 0_F and 1_F such that the following properties hold:

(FA1) addition is commutative, viz., $a + b = b + a$ for all $a, b \in F$;
(FA2) multiplication is commutative, viz., $a + b = b + a$ for all $a, b \in F$;
(FA3) addition is associative, viz., $a + (b + c) = (a + b) + c$ for all $a, b \in F$;
(FA4) multiplication is associative, viz., $a(bc) = (ab)c$ for all $a, b, c \in F$;
(FA5) addition and multiplication are distributive, viz., $a(b + c) = ab + ac$ and $(a + b)c = ac + bc$ for all $a, b, c \in F$;
(FA6) an additive identity 0_F exists, so that $a + 0_F = 0_F + a = a$ for all $a \in F$;
(FA7) a multiplicative identity 1_F exists, so that $a1_F = 1_F a = a$ for all $a \in F$;
(FA8) addition is invertible, viz., each $a \in F$ admits $-a \in F$ such that $a+(-a) = (-a) + a = 0_F$;
(FA9) multiplication is invertible, viz., each $a \in F \setminus \{0_F\}$ admits $a^{-1} \in F \setminus \{0_F\}$ such that $aa^{-1} = a^{-1}a = 1_F$.

Remark 1.2. If there is no danger of confusion, we simply write 0 and 1 for 0_F and 1_F, respectively. Moreover, given $n \in \mathbb{N}$ and $x \in F$, we write $n \cdot x$ to denote the n-fold sum $x + \cdots + x$.

Example 1.3. The set of real numbers $\mathbb{R}$ with usual addition and multiplication is a field. Similarly, the set of complex numbers $\mathbb{C}$ and the set of rational numbers $\mathbb{Q}$ with usual addition and multiplication are also fields. □

Proposition 1.4. *Let F be a field.*

(1) 0_F is the unique additive identity, and 1_F is the unique multiplicative identity.
(2) Additive inverses and multiplicative inverses are unique.
(3) $0_F \neq 1_F$.
(4) $a0_F = 0_F a = 0$ for all $a \in F$.
(5) $-(ab) = (-a)b = a(-b)$ for all $a, b \in F$.

Proof. Routine. □

Note that $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all share the same addition and multiplication operations. In what sense is $\mathbb{Q}$ a substructure of $\mathbb{R}$, or $\mathbb{C}$ a superstructure of $\mathbb{R}$? We formalize these notions below.

Definition 1.5. A *subfield* of a field F is a subset K of F such that

(1) $0_F, 1_F \in K$;

(2) $a + b \in K$ and $ab \in K$ whenever $a, b \in K$;
(3) $-a \in K$ and $a^{-1} \in K$ whenever $a, b \in K$.

Remark 1.6. It follows at once that K is a field with the same addition and multiplication operations as F.

Definition 1.7. A *field extension* of a field K is a field F that contains K as a subfield.

Example 1.8. $\mathbb{Q}$ has no proper subfield. Indeed, every subfield F of $\mathbb{Q}$ must contain all the integers and their multiplicative inverses. Therefore, $\frac{1}{n} \in F$ for all $n \in \mathbb{Z} \setminus \{0\}$. Moreover, $\frac{m}{n} = \sum_{i=1}^{m} \frac{1}{n}$ must be contained in F, whence $F \supseteq \mathbb{Q}$. It follows that $F = \mathbb{Q}$. □

Example 1.9. $\mathbb{R}$ is a field extension of $\mathbb{Q}$. $\mathbb{C}$ is a field extension of $\mathbb{R}$ or $\mathbb{Q}$. □

Example 1.10. Let d be a *squarefree integer*, viz., $d \neq 0$, $d \neq 1$, and the prime factorization of d does not repeat any prime. The subset

$$\mathbb{Q}[\sqrt{d}] = \{a + b\sqrt{d} : a, b \in \mathbb{Q}\}$$

of $\mathbb{C}$ is a subfield of $\mathbb{C}$, called a *quadratic field*. If $d > 0$, then $\mathbb{Q}[\sqrt{d}]$ is said to be a *real quadratic field*; if $d < 0$, then $\mathbb{Q}[\sqrt{d}]$ is said to be an *imaginary quadratic field*.

Example 1.11 (Modular arithmetic). So far, every example of a field we have considered was infinite. Let us now construct finite fields. For each $k \in \mathbb{Z}^+$, we define the following equivalence relation on $\mathbb{Z}$:

$$n \sim_k m \text{ if and only if } n - m \text{ is divisible by } k.$$

We denote by $\mathbb{Z}/k\mathbb{Z}$ the quotient set $\mathbb{Z}/\sim_k$ and define the addition and multiplication operations by setting

$$[n] + [m] = [n + m] \quad \text{and} \quad [n] \cdot [m] = [n \cdot m]$$

for all $n, m \in \mathbb{Z}$. It is not hard to check that these operations are well-defined.

Note that modular addition is always invertible, as $[k-n]$ is the additive inverse of $[n]$. Most integers, however, do not have multiplicative inverses, and so invertibility of modular multiplication is a trickier matter.

We claim that $\mathbb{Z}/k\mathbb{Z}$ is not a field if k is not a prime. To see this, we suppose that p is a prime number strictly less than k that divides k. We can then find $m \in \mathbb{N}$ such that $k = pm$. If $[pm] = [1]$ for some $n \in \mathbb{Z}$, then $pn - 1 = kn' = pmn'$ for some $n' \in \mathbb{Z}$, so that $p(n - mn') = 1$. But $p \geq 2$, and so p does not admit a multiplicative inverse in $\mathbb{Z}$. This is absurd, and we conclude that p has no multiplicative inverse in $\mathbb{Z}/k\mathbb{Z}$.

Let us now show that $\mathbb{Z}/k\mathbb{Z}$ is a field whenever k is prime. To this end, we make use of the following lemma:

Lemma 1.12 (Bézout's identity). *For all $n, m \in \mathbb{Z}$, we can find $a, b \in \mathbb{Z}$ such that*

$$an + bm = \gcd(n, m).$$

Proof of lemma. We first show that if

$$\{an + bm : a, b \in \mathbb{Z}\} = \{cd : c \in \mathbb{Z}\},$$

and $d > 0$, then then $d = \gcd(a, b)$. Since $m, n \in \{cd : c \in \mathbb{Z}\}$, we see that $d \mid n$ and $d \mid m$, and so d is a common divisor of n and m. If d' is an arbitrary common divisor of n and m, then $d' \mid an+bm$ for all $a, b \in \mathbb{Z}$. Since $d \in \{an+bm : a, b \in \mathbb{Z}\}$, we see that $d' \mid d$. It follows that $d = \gcd(m, n)$ as was to be shown.

What we have shown above implies that $\gcd(m, n) \in \{an + bm : a, b \in \mathbb{Z}\}$, and the lemma now follows. □

We now fix a nonzero integer n and apply the above lemma to find $m_1, m_2 \in \mathbb{Z}$ such that $m_1 n + m_2 k = \gcd(n, k)$. Since k is assumed to be prime, $\gcd(n, k) = 1$. We now observe that

$$[1] = [m_1 n + m_2 k] = [m_1][n] + [m_2 k] = [m_1][n] + [0] = [m_1][n],$$

whence $[m_1]$ is the multiplicative inverse of $[n]$. □

Note that $k \cdot [1] = [0]$ in $\mathbb{Z}/k\mathbb{Z}$, even though there is no integer n such that $n \cdot 1 = 0$ in $\mathbb{Z}$. This difference is captured by the following notion:

Definition 1.13. A field F is *of characteristic* k in case k is the smallest positive integer such that $k \cdot 1_F = 0$. If no such positive integer exists, then we say that F is *of characteristic zero*. We write $\text{char}(F)$ to denote the characteristic of F.

Example 1.14. $\mathbb{Q}$ is of characteristic zero. □

Example 1.15. $\mathbb{Z}/p\mathbb{Z}$ is of characteristic p. □

The above two examples cover all possible cases, as far as the computation of characteristics is concerned:

Proposition 1.16. *Every field of positive characteristic must be of prime characteristic.*

Proof. Let $k = \text{char}(F)$. We suppose for a contradiction that $k = mn$ for some integers $m, n \geq 2$. Since $m < k$ and $n < k$, we can find elements $a, b \in F$ such that $m \cdot a \neq 0$ and $n \cdot b \neq 0$, respectively. Now,

$$0 = mn \cdot ab = (m \cdot a)(n \cdot b).$$

But then

$$\begin{aligned} 0 &= 0(m \cdot a)^{-1}(n \cdot b)^{-1} = (m \cdot a)(n \cdot b)(m \cdot a)^{-1}(n \cdot b)^{-1} \\ &= (m \cdot a)(m \cdot a)^{-1}(n \cdot b)(n \cdot b)^{-1} = 1, \end{aligned}$$

which is absurd. We conclude that F is of prime characteristic. □

The above proposition shows that the subset

$$P = \{0_F, 1_F, 2 \cdot 1_F, \ldots, (\text{char}(F) - 1) \cdot 1_F\}$$

of a field F of positive characteristic forms a subfield of F. In other words, a field of positive characteristic contains a copy of $\mathbb{Z}/\text{char}(F)\mathbb{Z}$. To formalize this, we need to introduce the notion of *field isomorphism*.

Definition 1.17. A *field homomorphism*, or a *morphism of fields*, is a function $\varphi : F \to K$ between two fields F and K that satisfies the following properties:

(1) $\varphi(a + b) = \varphi(a) + \varphi(b)$ for all $a, b \in F$;
(2) $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in F$;
(3) $\varphi(0_F) = 0_F$;
(4) $\varphi(1_F) = 1_F$.

Proposition 1.18. *The collection of all fields with field homomorphisms forms a category, denoted by* **Field***. Moreover:*

(1) monomorphisms in **Field** *are injective field homomorphisms;*
(2) epimorphisms in **Field** *are surjective field homomorphisms;*
(3) isomorphisms in **Field***, called* field isomorphisms*, are bijective field homomorphisms;2*
(4) every morphism in **Field** *is a monomorphism;*
(5) every epimorphism in **Field** *is an isomorphism.*
(6) Given a field homomorphism $\varphi : F \to K$*, the image* $\operatorname{im}\varphi$ *is a subfield of* K *that is isomorphic to* F*.*

Proof. Checking the category axioms is trivial.

(1) Let $f : F \to K$ be a function. Suppose that f is a monomorphism in **Field**. This, in particular, implies that f is a monomorphism in **Set**, whence Example 17.8 implies that f is an injective function. The assumption also implies that f is a field homomorphism.

Conversely, suppose that f is an injective field homomorphism. Let $g_1, g_2 : S \to F$ be field homomorphisms such that $f \circ g_1 = f \circ g_2$. This, in particular, implies that $f \circ g_1 = f \circ g_2$ in **Set**, and so Example 17.8 implies that $g_1 = g_2$.

(2) Let $f : F \to K$ be a function. Suppose that f is an epimorphism in **Field**. This, in particular, implies that f is an epimorphism in **Set**, whence Example 17.8 implies that f is a surjective function. The assumption also implies that f is a field homomorphism.

Conversely, suppose that f is a surjective field homomorphism. Let $h_1, h_2 : K \to S$ be field homomorphisms such that $h_1 \circ f = h_2 \circ f$. This, in particular, implies that $h_1 \circ f = h_2 \circ f$ in **Set**, and so Example 17.8 implies that $h_1 = h_2$.

(3) If $\varphi : F \to K$ is a bijective field homomorphism, then $\varphi^{-1}(0_K) = 0_F$ and $\varphi^{-1}(1_K) = 1_F$. Moreover, given $a, b \in K$, we set $c = \varphi^{-1}(a)$ and $d = \varphi^{-1}(d)$. We then observe that

$$\varphi^{-1}(a+b) = \varphi^{-1}(\varphi(c)+\varphi(d)) = \varphi^{-1}(\varphi(c+d)) = c+d = \varphi^{-1}(a)+\varphi^{-1}(b)$$

and that

$$\varphi^{-1}(ab) = \varphi^{-1}(\varphi(c)\varphi(d)) = \varphi^{-1}(\varphi(cd)) = cd = \varphi^{-1}(a)\varphi^{-1}(b).$$

Therefore, φ is an isomorphism in **Field**. The converse is trivial.

(4) Let $\varphi : F \to K$ be a field homomorphism. We define the *kernel* of φ to be the set

$$\ker\varphi = \{x \in F : \varphi(x) = 0_K\}.$$

We first show that φ is injective if and only if If φ is injective, then the kernel is clearly trivial. Conversely, we suppose that the kernel is trivial and fix $a, b \in F$ such that $\varphi(a) = \varphi(b)$. Note that $\varphi(a-b) = 0_K$, so that $a - b \in \ker\varphi$. It follows that $a = b$, whence φ is injective.

Let us now define an *ideal* of F to be a subset I of F such that

- $a \in F$ and $x \in I$ imply $ax \in I$;
- $x, y \in I$ implies $x + y \in I$.

We show that F has only two ideals: $\{0_F\}$ and F. Indeed, if x is a nonzero element of an ideal I, then $1_F = x^{-1}x \in I$, and so $a = a1_F \in I$ for all $a \in F$.

Observe that $\ker\varphi$ is an ideal of F. Indeed, if $a \in F$ and $x \in \ker\varphi$, then

$$\varphi(ax) = \varphi(a)\varphi(x) = \varphi(a)0_K = 0_K,$$

so that $ax \in \ker \varphi$. Moreover, if $x, y \in \ker \varphi$, then

$$\varphi(x+y) = \varphi(x) + \varphi(y) = 0_K + 0_K = 0_K,$$

and so $x + y \in \ker \varphi$.

Now, $\varphi(1_F) = 1_K \neq 0_K$, and so $1_F \notin \ker \varphi$. Since $\ker \varphi$ is an ideal, it follows that $\ker \varphi = \{0_F\}$, whence φ is injective. The desired result now follows from (1).

(5) This is a direct consequence of (1), (2), (3), and (4).

(6) Let $\varphi : F \to K$ be a field homomorphism. Evidently, $0_K = \varphi(0_F)$ and $1_K = \varphi(1_F)$. We fix $a, b \in \operatorname{im} \varphi$ and find $c, d \in F$ such that $a = \varphi(c)$ and $b = \varphi(d)$. Observe that

$$a + b = \varphi(c) + \varphi(d) = \varphi(c + d) \in \operatorname{im} \varphi.$$

Similarly, $ab \in \operatorname{im} \varphi$. We also note that

$$1_K = \varphi(c - c) = \varphi(c)\varphi(-c) = a\varphi(-c),$$

whence $-a = \varphi(-c) \in \operatorname{im} \varphi$. Similarly, $a^{-1} \in \operatorname{im} \varphi$. We conclude that $\operatorname{im} \varphi$ is a subfield of K.

It now follows from (5) that $F \cong \operatorname{im} \varphi$. □

As a simple application, we establish the following property of characteristics:

Corollary 1.19. *If there exists a field homomorphism $\varphi : F \to K$, then* $\operatorname{char}(F) = \operatorname{char}(K)$.

Proof. Proposition 1.18 implies that φ is an injective field homomorphism.

If $\operatorname{char}(F) = 0$, then $n \cdot 1_F \neq 0_F$ for all $n \geq 1$. Since φ is injective,

$$n \cdot 1_K = n \cdot \varphi(1_F) = \varphi(n \cdot 1_F)$$

is never 0_K for any $n \geq 1$. We conclude that $\operatorname{char}(K) = 0$.

If F is of positive characteristic, then an analogous argument implies that

$$n \cdot 1_K = n \cdot \varphi(1_F) = \varphi(n \cdot 1_F)$$

is nonzero for all $1 \leq n \leq \operatorname{char}(F) - 1$ and is zero when $n = \operatorname{char}(F)$. It follows that $\operatorname{char}(F) = \operatorname{char}(K)$. □

What the above corollary tells us is that it is necessary to have matching characteristics in order for field homomorphisms $\varphi : F \to K$ to exist. We now show that, if F is either $\mathbb{Q}$ or $\mathbb{Z}/p\mathbb{Z}$, then it is *sufficient* to have matching characteristics to ensure the existence of field homomorphisms. To lay out the proper context for this result, we introduce the notion of the *prime subfield* of a field, i.e., the subfield generated by 1.

Definition 1.20. Let F be a field and let S be a subset of F. The *subfield of F generated by S* is the smallest subfield of F that contains S, viz., the intersection of all subfields of F that contains S. If $S = \{1\}$, then the subfield generated by S is referred to as the *prime subfield* of F.

The promised sufficiency result is established below:

Theorem 1.21. *Let F be a field. If* $\operatorname{char}(F) = 0$, *then the prime subfield of F is isomorphic to $\mathbb{Q}$. If* $\operatorname{char}(F) = p$ *for some prime p, then the prime subfield of F is isomorphic to $\mathbb{Z}/p\mathbb{Z}$.*

Proof. Let F be a field and let P be the prime subfield of F. If $\mathrm{char}(F) = 0$, then the mapping $(n/m) = n \cdot (m \cdot 1_F)^{-1}$ is a field homomorphism from $\mathbb{Q}$ into F. Since $1_F \in \operatorname{im}\varphi$, we see that $\operatorname{im}\varphi$ is a field extension of P. now, $\operatorname{im}\varphi \cong \mathbb{Q}$ by Proposition 1.18(6). Example 1.8 shows that $\mathbb{Q}$ has no proper subfield, and so $P = \operatorname{im}\varphi \cong \mathbb{Q}$.

If $\mathrm{char}(F) = p$, then the mapping $\phi(n) = n \cdot 1_F$ is a field homomorphism from $\mathbb{Z}/p\mathbb{Z}$ into F. Analogously as above, we can deduce that $P = \operatorname{im}\phi \cong \mathbb{Z}/p\mathbb{Z}$. □

2. Vector Spaces

Let us now consider generalizations of the Euclidean n-space $\mathbb{R}^n$ and its complex counterpart $\mathbb{C}^n$.

Definition 2.1. Let F be a field. A *vector space over* F, or an F*-vector space*, is a set V equipped with *vector addition* $V \times V \to V$ and *scalar multiplication* $F \times V \to V$ that satisfy the following properties:

(1) vector addition is commutative, viz., $v + w = w + v$ for all $v, w \in V$;
(2) vector addition is associative, viz., $(v+w)+x = v+(w+x)$ for all $v, w, x \in V$;
(3) the additive identity 0_V exists, so that $v + 0_V = 0_V + v = v$ for all $v \in V$;
(4) vector addition is invertible, viz., each $v \in V$ admits $-v \in V$ such that $v + (-v) = (-v) + v = 0_V$;
(5) scalar multiplication is associative, viz., $a(bv) = (ab)v$ for all $a, b \in V$;
(6) scalar multiplication is distributive, viz., $a(v + w) = av + aw$ and $(a + b)v = av + bv$ for all $a, b \in F$ and $v, w \in V$;
(7) $1_F v = v$ for all $v \in V$;
(8) $0_F v = 0_V$ for all $v \in V$;
(9) $-1_F v = -v$ for all $v \in V$.

Example 2.2. If we have a field F and a subfield K, then F is a vector space over K. □

Example 2.3. Given a field, we define

$$F^n = \{(v^1, \ldots, v^n) : v^1, \ldots, v^n \in F\}$$

and endow it with coordinate-wise vector addition

$$(v^1, \ldots, v^n) + (w^1, \ldots, w^n) = (v^1 + w^1, \ldots, v^n + w^n)$$

for all $v^1, \ldots, v^n, w^1, \ldots, w^n \in F$ and scalar multiplication

$$a(v^1, \ldots, v^n) = (av^1, \ldots, av^n)$$

for all $a, v^1, \ldots, v^n \in F$. F^n is easily seen to be a vector space over F. □

Example 2.4. We generalize the above example by considering infinite copies of a field F. Given an index set I, we consider the I-fold cartesian product

$$F^I = \prod_{\alpha \in I} F = \{(v^\alpha)_{\alpha \in I} : v^\alpha \in F\}$$

and endow it with coordinate-wise vector addition

$$a(v^\alpha)_{\alpha \in I} = (av^\alpha)_{\alpha \in I}.$$

F^I is easily seen to be a vector space over F. □

Example 2.5. Although the construction of F^I seems completely natural, we shall see in due course that F^I is not a very useful object. Here we construct a substructure of F^I that will serves as the model vector space in developing the theory of linear algebra.

Given a field F and an index set I, we define $F^{\oplus I}$ to be the subset of F^I consisting of vectors $(v^\alpha)_{\alpha \in I}$ such that $v^\alpha = 0$ for all but finitely many indices α. If we endow $F^{\oplus I}$ with coordinate-wise vector addition and scalar multiplication, then $F^{\oplus I}$ becomes a vector space over F. Indeed, the key fact to observe is that the coordinate-wise vector sum of two elements of $F^{\oplus I}$ still belongs to $F^{\oplus I}$. □

The above construction motivates the following definition of substructures of a vector space:

Definition 2.6. Let V be a vector space over a field F. A *linear subspace* of V is a subset W of V such that

(1) $0_V \in W$;
(2) if $v \in W$, then $av \in W$ for all $a \in F$;
(3) if $v, w \in W$, then $v + w \in W$.

Remark 2.7. It follows at once that W is a vector space with the same vector addition and scalar multiplication operations as V.

We also introduce structure-preserving maps for the category of vector spaces.

Definition 2.8. Let F be a field. A *linear transformation* is a function $T : V \to W$ between two vector spaces V and W over F that satisfies the following properties:

(1) $T(v + w) = T(v) + T(w)$ for all $v, w \in V$;
(2) $T(av) = aT(v)$ for each $a \in F$ and every $v \in V$;
(3) $T(0_V) = 0_W$.

A linear transformation is also referred to as a *linear map*,' an *F-linear map*, a *linear mapping*, an *F-linear mapping*, a *linear operator*, or a *morphism of vector spaces.* A bijective linear transformation is said to be a *linear isomorphism*, or simple an *isomorphism.* We say that two vector spaces V and W over F are *isomorphic*, or *linearly isomorphic*, if there exists a linear isomorphism $V \to W$.

Remark 2.9. We often drop the parentheses and write Tv to denote $T(v)$.

Proposition 2.10. *The collection of all vector spaces over a fixed field F with linear transformations forms a category, denoted by $F-$**Vect**. Moreover:*

*(1) monomorphisms in $F-$**Vect** are injective linear transformations;*
*(2) epimorphisms in $F-$**Vect** are surjective linear transformations;*
*(3) isomorphisms in $F-$**Vect**, called* linear isomorphisms*, are bijective linear transformations;*
(4) Given an injective linear transformation $T : V \to W$, we have the isomorphism $\operatorname{im} T \cong V$. (See Definition 5.1)

Proof. A routine modification of the proof of Proposition 1.18 works. □

3. Bases, Products, Sums, and Free Vector Spaces

We shall show that every vector space is isomorphic $F^{\oplus I}$ for some index set I. To this end, we single out the basic building blocks of vector spaces.

Definition 3.1. Let V be a vector space over a field F. A *linear combination* of $v_1, \dots, v_n \in V$ is a vector in V of the form

$$a_1 v_1 + \cdots + a_n v_n, \tag{3.2}$$

where $a_1, \dots, a_n \in V$. We say that $\{v_1, \dots, v_n\}$ is *linearly independent* in case (3.2) equals zero if and only if $a_1 = \cdots = a_n = 0$. The *span* of $\{v_1, \dots, v_n\}$ is the set

$$\operatorname{span}\{v_1, \dots, v_n\} = \{a_1 v_1 + \ldots + a_n v_n : a_1, \dots, a_n \in F\}$$

of all linear combinations of $v_1, \dots, v_n$.

Definition 3.3. Let V be a vector space over a field F, I an index set, and $\{v_\alpha\}_{\alpha \in I} \subseteq V$. We say that $\{v_\alpha\}_{\alpha \in I}$ is *linearly independent* if every finite subset thereof is linearly independent. a set that is not linearly independent is said to be *linearly dependent*. The *span* of $\{v^\alpha\}_{\alpha \in I}$ is the union

$$\operatorname{span}\{v_\alpha\}_{\alpha \in I} = \bigcup_{\substack{A \subseteq \{v_\alpha\}_{\alpha \in I} \\ A \text{ is finite}}} \operatorname{span} A.$$

$\{v^\alpha\}_{\alpha \in I}$ is said to be a *basis of* V if $\operatorname{span}\{v_\alpha\}_{\alpha \in I} = V$ and $\{v_\alpha\}_{\alpha \in I}$ is linearly independent.

Remark 3.4. We note that the span of any subset of a vector space V is a linear subspace of V.

Remark 3.5. We insist on taking *finite* linear combinations, as infinite sums do not make sense in general vector spaces.

Example 3.6. Consider $F^{\oplus I}$ that was constructed in Example 2.5. We define the *standard coordinate vectors* to be the set $\{e_\alpha\}_{\alpha \in I}$ of vectors

$$e_\alpha(\beta) = \begin{cases} 1 & \text{if } \alpha = \beta; \\ 0 & \text{otherwise.} \end{cases}$$

$\{e_\alpha\}_{\alpha \in I}$ is easily seen to be a basis of $F^{\oplus I}$. □

The most fundamental property of a basis is that every vector admits a unique representation as a finite linear combination of the basis elements.

Proposition 3.7. *Let V be a vector space over F. If V admits a basis $\{v_\alpha\}_{\alpha \in I}$, then there exists a unique function $a : V \setminus \{0\} \to \prod_{\alpha \in I} F$ such that, for each $v \in V \setminus \{0\}$,*

(1) $a(v)(\alpha) = 0$ for all but finitely many $\alpha \in I$, and
(2) $v = \sum_{\alpha \in I} a(v)(\alpha) v_\alpha$

Remark 3.8. The sum in (2) always makes sense, as (1) guarantees that the sum is finite.

Proof. The existence of such a function follows at once from the fact that $\{v_\alpha\}_{\alpha \in I}$ is a basis of V. To prove uniqueness, we suppose there are two such functions, a and b. Fix a nonzero vector v in V, so that

$$v = \sum_{\alpha \in I} a(v)(\alpha) v_\alpha = \sum_{\alpha \in I} b(v)(\alpha) v_\alpha.$$

This, in particular, implies that

$$0 = v - v = \sum_{\alpha \in I} (a(v)(\alpha) - b(v)(\alpha))\, v_\alpha.$$

Since $\{v_\alpha\}_{\alpha \in I}$ is a linearly independent subset of V, we conclude that $a(v)(\alpha) - b(v)(\alpha) = 0$ for all $\alpha \in I$, whence $a(v) = b(v)$. v was chosen arbitrarily, and so $a = b$. □

Theorem 3.9. *Let V be a vector space over a field F. Every set of linearly independent vectors in V is contained in a basis of V.*

The proof below is an archetypal "zornification" argument: construct a small extension by hand, use Zorn's lemma to create a maximal object, and verify that this maximal object is the desired one.

Proof. Fix a set A of linearly independent vectors in V. If A is a *maximal* linearly-independent set, viz., every superset B of A is linearly dependent in V, then A is a basis of V. Indeed, if there exists a vector $v \in V$ such that no finite linear combination in A equals v, then $A \cup \{v\}$ is a linearly-independent superset of A, contradicting the maximality condition. We therefore assume that A is non-maximal.

Let $(\mathcal{A}, \preceq)$ be the collection of all linearly-independent supersets of A in V, ordered by the set-inclusion relation: $B \preceq C$ if and only if $B \subseteq C$. If $\{B_\alpha : \alpha \in I\}$ is a chain in $\mathcal{A}$, then $\bigcup_{\alpha \in I} B_\alpha$ is an upper bound of $\{B_\alpha : \alpha \in I\}$ in $\mathcal{A}$. It now follows from Zorn's lemma (Theorem 18.21) that there exists a maximal element B of $\mathcal{A}$, which, by construction, is a maximal linearly-independent subset of V. □

Bases provide an extremely important isomorphic invariant in $F-$**Vect**.

Definition 3.10. The *dimension* of a vector space V over a field F is the cardinal number corresponding to an arbitrary basis of V.

Theorem 3.11. *Every basis of a fixed vector space V over a field F is of the same cardinality. In particular, the dimension of V is well-defined.*

Corollary 3.12. *Dimension is an isomorphic invariant in $F-$**Vect**.*

Proof of theorem and corollary. We assume for now that V has a finite spanning set. It follows at once from the claim below that all bases of V must be of the same cardinality.

Claim. *If $\{v_1, \ldots, v_n\}$ is a linearly independent subset of V, and if $\{s_1, \ldots, s_m\}$ is a spanning set of V, then $n \leq m$.*

Proof of claim. Consider the list

$$s_1, \ldots, s_m \mid v_1, \ldots, v_n;$$

the front list constitutes a spanning set of V, and the back list constitutes a set of linearly independent vectors in V. Now, $\{s_1, \ldots, s_n\}$ spans V, and so v_1 is a linear combination of some vectors in $\{s_1, \ldots, s_n\}$. Therefore, there exists an s_i such that $\{v_1, s_1, \ldots, s_n\} \smallsetminus \{s_i\}$ is still a spanning set of V. By relabeling if necessary, we assume without loss of generality that $i = 1$ and construct a new list:

$$v_1, s_2, \ldots, s_m \mid v_2, \ldots, v_n.$$

Note that the front list still constitutes a spanning set of V, and that the back list constitutes a set of linearly independent vectors in V.

We now suppose inductively that we have the list

$$v_1, \ldots, v_k, s_{k+1}, \ldots, s_m \mid v_{k+1}, \ldots, v_n$$

satisfying the properties specified above. The set $\{v_1, \ldots, v_k, s_{k+1}, \ldots, s_m\}$ spans V, and so v_{k+1} is a linear combination of some elements thereof. Since $\{v_1, \ldots, v_{k+1}\}$ is linearly independent, no linear combination of $v_1, \ldots, v_k$ is v_{k+1}. This implies that any subset of $\{v_1, \ldots, v_k, s_{k+1}, \ldots, s_m\}$ whose span includes v_{k+1} must contain at least one s_i for some $k+1 \leq i \leq m$. By relabeling if necessary, we assume without loss of generality that $i = k+1$ and construct a new list:

$$v_1, \ldots, v_{k+1}, s_{k+2}, \ldots, s_m \mid v_{k+2}, \ldots, v_n.$$

It follows that the front list constitutes a spanning set of V, and that the back list constitutes a set of linearly independent vectors in V.

Now, if $m < n$, then it follows from induction that the list

$$v_1, \ldots, v_m \mid v_{m+1}, \ldots, v_n$$

satisfies the properties prescribed above. But this is absurd: $\{v_1, \ldots, v_m, v_{m+1}\}$ is a set of linearly independent vectors in V, and so the span of $\{v_1, \ldots, v_m\}$ cannot include v_{m+1}. We conclude that $m \geq n$. □

We now assume that every spanning set of V is infinite. Suppose that $\{v_\alpha\}_{\alpha \in I}$ and $\{w_\beta\}_{\beta \in J}$ are two bases of V, both of which must be infinite. For each $\alpha \in I$, let J_α be the finite subset of J such that v_α is a linear combination of the elements of $\{w_\beta\}_{\beta \in J_\alpha}$. Proposition 3.7 guarantees that J_α is well-defined.

We note that

$$J = \bigcup_{\alpha \in I} J_\alpha. \tag{3.13}$$

If not, then we can find $\beta_0 \in J$ such that $\{w_\beta\}_{\beta \in J \smallsetminus \{\beta_0\}}$ is still a basis of V. This, in particular, implies that w_{β_0} is a linear combination of some elements of $\{w_\beta\}_{\beta \in J \smallsetminus \{\beta_0\}}$, whence $\{w_\beta\}_{\beta \in J}$ is no longer a linearly independent subset of V. This is evidently absurd.

Observe that $|J_\alpha| \leq \aleph_0$ for each α, where $\aleph_0$ is the first infinite cardinal. We now apply cardinal arithmetic (Definition 18.31 and the discussion that follows) to (3.13) to conclude that

$$|J| \leq \aleph_0 |I| = |I|.$$

A similar argument now shows that

$$|I| \leq |J|,$$

whence it follows from the Cantor–Bernstein theorem (Theorem 18.25) that

$$|I| = |J|,$$

as was to be shown. This proves the theorem.

It remains to show that dimension is an isomorphic invariant in $F-$**Vect**. To this end, we suppose that $T : V \to W$ is a linear isomorphism of two F-vector spaces V and W. We invoke Theorem 3.9 to find a basis $\{v_\alpha\}_{\alpha \in I}$ of V. Set $w_\alpha = Tv_\alpha$ for each $\alpha \in I$. It follows from the linearity of T that $\{w_\alpha\}_{\alpha \in I}$ is a linearly independent subset of W. Since T is bijective, we conclude that $\{w_\alpha\}_{\alpha \in I}$ is a spanning set of

W, whence $\{w_\alpha\}_{\alpha\in I}$ is a basis of W with the same cardinality as the basis $\{v_\alpha\}_{\alpha\in I}$ of V. □

Example 3.14. Example 3.6 implies that the vector space $F^{\oplus I}$ constructed in Example 2.5 is of dimension $|I|$, where $|I|$ denotes the cardinality of the index set I.

We finally show that dimension provides a complete isomorphic classification of vector spaces. Indeed, if V is of dimension κ, then V is isomorphic to $F^{\oplus I}$ for some index set I of cardinality κ.

Theorem 3.15. *Every vector space of the same dimension is linearly isomorphic.*

The proof employs the standard tactic of constructing a morphism by mapping generators to generators; we shall see additional examples of this throughout these notes.

Proof. Let V and W be two vector spaces of the same dimension. By Theorem 3.9, we can find bases $\{v_\alpha\}_{\alpha\in I}$ and $\{w_\beta\}_{\beta\in J}$ of V and W, respectively. Theorem 3.11 implies that $|I| = |J|$, whence there exists a bijection $f : I \to J$.

Let us now define a mapping $T : V \to W$ by setting $Tv_\alpha = w_{f(\alpha)}$ for each $\alpha \in I$ and extending linearly, i.e., setting

$$T\left(\sum_{n=1}^N a_n v_{\alpha_n}\right) = \sum_{n=1}^N a_n T v_{\alpha_n}$$

for all $\alpha_1, \ldots, \alpha_n \in I$ and $a_1, \ldots, a_n \in F$. It follows from the uniqueness of basis expansions (Proposition 3.7), as well as the fact that $\{v_\alpha\}_{\alpha\in I}$ and $\{w_\beta\}_{\beta\in J}$ are bases of V and W, respectively, that T is a linear isomorphism. □

In light of the isomorphic classification we have just carried out in Theorem 3.15, it is natural to seek for a construction of a "generic" F-vector space of dimension κ by piecing together κ copies of F. We make rigorous the notion of piecing together different vector spaces by mimicking the constructions of F^I and $F^{\oplus I}$:

Definition 3.16. The *direct product* of a family $(V_\alpha)_{\alpha\in I}$ of vector spaces over a field F is the cartesian product $\prod_{\alpha\in I} V_\alpha$ endowed with coordinate-wise vector addition

$$(v^\alpha)_{\alpha\in I} + (w^\alpha)_{\alpha\in I} = (v^\alpha + w^\alpha)_{\alpha\in I}$$

and scalar multiplication

$$a(v^\alpha)_{\alpha\in I} = (av^\alpha)_{\alpha\in I}.$$

The *direct sum* of $(V_\alpha)_{\alpha\in I}$ is the subspace $\bigoplus_{\alpha\in I} V_\alpha$ of $\prod_{\alpha\in I} V_\alpha$ consisting of vectors $(v^\alpha)_{\alpha\in I}$ such that $v^\alpha = 0$ for all but finitely many indices α.

Remark 3.17. It follows at once that finite direct products agree with finite direct products.

We show that these two constructions correspond to the categorical product and the categorical coproduct.

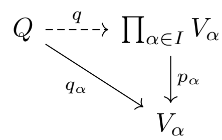
Proposition 3.18. *Let $(V_\alpha)_{\alpha\in I}$ be a family of vector spaces over a field F. The pair $(\prod_{\alpha\in I} V_\alpha, \{p_\alpha\}_{\alpha\in I})$ of the direct product $\prod_{\alpha\in I} V_\alpha$ and the collection $\{p_\alpha\}_{\alpha\in I}$ of the standard coordinate functions $p_\alpha : \prod_{\alpha\in I} V_\alpha \to V_\alpha$ is a product in $F-$**Vect**. The pair $(\bigoplus_{\alpha\in I} V_\alpha, \{\iota_\alpha\}_{\alpha\in I})$ of the direct sum $\bigoplus_{\alpha\in I} V_\alpha$ and the collection $\{\iota_\alpha\}_{\alpha\in I}$*

with the canonical injection maps $\iota_\alpha : V_\alpha \to \bigoplus_{\alpha\in I} V_\alpha$ *that sends* $v \in V_\alpha$ *to* $(v^\beta)_{\beta\in I}$ *defined by setting*

$$v^\beta = \begin{cases} v & \text{if } \beta = \alpha; \\ 0 & \text{otherwise}; \end{cases}$$

is a coproduct in $F-\mathbf{Vect}$.

Proof. Let $(Q, \{q_\alpha\}_{\alpha\in I})$ be a pair consisting of a vector space Q and a family $\{q_\alpha\}$ of F-linear mappings $q_\alpha : Q \to V_\alpha$. The map $q : Q \to \prod_{\alpha\in I} V_\alpha$ that sends v to $(q_\alpha(v))_{\alpha\in I}$ is evidently the unique F-linear mapping such that the diagram

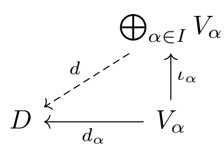


commutes for all $\alpha \in I$. It follows that $(\prod_{\alpha\in I} V_\alpha, \{p_\alpha\}_{\alpha\in I})$ is a product in $F-\mathbf{Vect}$.

We now let $(D, \{d_\alpha\}_{\alpha\in I})$ be a pair consisting of a vector space D and a family $\{d_\alpha\}_{\alpha\in I}$ of F-linear mappings $d_\alpha : V_\alpha \to D$. We define the map $d : \bigoplus_{\alpha\in I} V_\alpha \to D$ by setting $d(\iota_\alpha(v)) = d_\alpha(v)$ for each $\alpha \in I$ and every $v \in V_\alpha$ and extending d linearly. This is well-defined, as

$$\operatorname{span}\left(\bigcup_{\alpha\in I}\{\iota_\alpha(v) : v \in V_\alpha\}\right) = \bigoplus_{\alpha\in I} V_\alpha.$$

By construction, d is evidently the unique F-linear mapping such that the diagram



commutes for all $\alpha \in I$. It follows that $(\bigoplus_{\alpha\in I} V_\alpha, \{\iota_\alpha\}_{\alpha\in I})$ is a coproduct in $F-\mathbf{Vect}$. □

Example 3.19. It now follows at once that $F^I = \prod_{\alpha\in I} F$ and that $F^{\oplus I} = \bigoplus_{\alpha\in I} F$, where F is considered as a one-dimensional vector space over F. □

We now show that infinite direct products are larger than infinite direct sums. Compare the result below with Proposition 3.26, which shows that finite direct sums agree with finite direct products.

Theorem 3.20. *For each index set* I *and a family* $\{V_\alpha\}_{\alpha\in I}$ *of vector spaces over a field* F*, we have the following inequality of cardinal numbers:*

$$\dim\left(\bigoplus_{\alpha\in I} V_\alpha\right) \leq \dim\left(\prod_{\alpha\in I} V_\alpha\right).$$

The equality holds if and only if $|I| < \infty$.

Proof. By the isomorphic classification of vector spaces (Theorem 3.15), it evidently suffices to compare the dimensions of $F^{\oplus I}$ and F^I, constructed in Example 2.4 and Example 2.5, respectively. If $|I| < \infty$, then $F^I = F^{\oplus I}$, and so

$$\dim F^{\oplus I} = \dim F^I.$$

We thus assume that I is an infinite index set.

We shall show that

$$\dim_F F^{\oplus I} = \dim_K K^{\oplus I} < \dim_K K^I \leq \dim_F F^I, \tag{3.21}$$

where K is the prime subfield of F. To this end, we recall from Theorem 1.21 that K is always countable.

We have already seen that $\dim_F F^{\oplus I} = |I| = \dim_K K^{\oplus I}$. As for the third inequality, we fix a finite set $\{v_n\}_{n=1}^N$ of linearly independent vectors in K^I and suppose that

$$\sum_{n=1}^N a_n v_n = 0$$

for some $a_1, \ldots, a_N \in F$. We consider F as a vector space over K and invoke Theorem 3.9 to find a basis $\{f_\delta\}_{\delta \in J}$ of F. For each n, we take the basis expansion

$$a_n = \sum_{\delta \in J} b_{n,\delta} f_\delta$$

of a_n. Since each of these sums is finite, we swap the sums to conclude that

$$0 = \sum_{n=1}^N a_n v_n = \sum_{n=1}^N \sum_{\delta \in J} b_{n,\delta} f_\delta v_n = \sum_{\delta \in J} \left(\sum_{n=1}^N b_{n,\delta} v_n \right) f_\delta.$$

We let $c_\delta(\beta) = b_{n,\delta} v_n(\beta)$ for each $\beta \in I$, which is a scalar in K. The above identity now implies, for each fixed $\beta \in I$, that

$$0 = \sum_{\delta \in J} c_\delta(\beta) f_\delta,$$

whence $c_\delta(\beta) = 0$ for all $\delta \in J$. It follows that $\{v_n\}_{n=1}^N$ is linearly independent in F^I. This, in particular, implies that any basis of K^I is linearly independent in F^I, whence $\dim_K K^I \leq \dim_F F^I$.

It remains to show that $\dim_K K^{\oplus I} < \dim_K K^I$. To this end, we recall from Theorem 3.15 that $\dim V = \dim W$ if and only if there exists a linear isomorphism $T : V \to W$. Linear isomorphisms are bijections, and so vector spaces of the same dimension must be of the same cardinality. Therefore, if we showed that $|K^{\oplus I}| < |K^I|$, then the desired result would follow.

For each $J \subseteq I$, we define

$$\mathcal{F}_J = \{f \in K^{\oplus I} : \operatorname{supp} f = J\},$$

so that $\{\mathcal{F}_J\}_{J \in \mathcal{P}(I)}$ is a pairwise-disjoint collection. Observe that

$$K^{\oplus I} = \bigcup_{\substack{J \in \mathcal{P}(I) \\ |J| < \infty}} \mathcal{F}_J.$$

Since the above union is disjoint, we apply cardinal arithmetic (Definition 18.31 and the discussion that follows) to conclude that

$$\left| K^{\oplus I} \right| = \sum_{\substack{J \in \mathcal{P}(I) \\ |J| < \infty}} |\mathcal{F}_J|$$

For each finite subset J of I, there is an obvious injection from $\mathcal{F}_J$ into the $|J|$-fold cartesian product $\prod_{\alpha \in J} K$ of K. Since J is finite and K is countable, we see that

$$|\mathcal{F}_J| \leq \left| \prod_{\alpha \in J} K \right| = |K| = \aleph_0,$$

where $\aleph_0$ is the first infinite cardinal. It now follows that

$$\left|K^{\oplus I}\right| = \sum_{\substack{J \in \mathcal{P}(I) \\ |J| < \infty}} |\mathcal{F}_J| \leq \sum_{\substack{J \in \mathcal{P}(I) \\ |J| < \infty}} \aleph_0.$$

We now let I_n be the collection of n-element subsets of I, so that

$$\bigcup_{n=1}^{\infty} I_n = \{J \in \mathcal{P}(I) : |J| < \infty\}.$$

For each $n \in \mathbb{N}$, there is an obvious injection from I_n into the n-fold cartesian product $\prod_{k=1}^n I$ of I. Since $|I|$ is infinite, we have the cardinal identity $|\prod_{k=1}^n I| = |I|$, and so

$$|\{J \in \mathcal{P}(I) : |J| < \infty\}| = \left| \bigcup_{n=1}^{\infty} I_n \right| = \sum_{n=1}^{\infty} |I| = \aleph_0 |I| = \max(\aleph_0, |I|) = |I|.$$

It follows that

$$\left|K^{\oplus I}\right| \leq \sum_{\substack{J \in \mathcal{P}(I) \\ |J| < \infty}} \aleph_0 \leq \aleph_0 |I| = \max(\aleph_0, |I|) = |I|.$$

Since there is an obvious injection from I into $K^{\oplus I}$, we have the reverse inequality $|I| \leq |K^{\oplus I}|$. We now conclude from the Cantor–Bernstein theorem (Theorem 18.25) that

$$\left|K^{\oplus I}\right| = |K|.$$

On the other hand, there is an obvious injection from $\mathbb{Z}_2^I$ to K^I, and so

$$\left|K^I\right| \geq |\mathbb{Z}_2^I| = 2^{|I|} = |\mathcal{P}(I)|.$$

It now follows from Cantor's theorem (Theorem 18.26) that

$$\left|K^I\right| \geq |\mathcal{P}(I)| > |I| = \left|K^{\oplus I}\right|,$$

as was to be shown. □

If all vector spaces in question are contained in a larger vector space, then there is an intuitive way of constructing the direct sum.

Definition 3.22. Let V be a vector space over F. The *Minkowski sum* of two subsets A and B of V is the set

$$A + B = \{v + w : v \in A \text{ and } w \in B\};$$

The Minkowski sum of a finite collection of subsets of V is defined analogously. The Minkowski sum of an arbitrary collection $\{A_\alpha\}_{\alpha \in I}$ of subsets of V is the set

$$\sum_{\alpha \in I} A_\alpha = \bigcup_{\substack{J \subseteq I \\ |J| < \infty}} \sum_{\alpha \in J} A_\alpha,$$

where $\sum_{\alpha \in J} A_\alpha$ denotes the finite Minkowski sum.

Remark 3.23. We remark that the Minkowski sum of linear subspaces of a vector space V is always a linear subspace of V.

Definition 3.24. A vector space V over a field F is said to be the *internal direct sum* of a family $\{A_\alpha\}_{\alpha\in I}$ of linear subspaces of V if V is the Minkowski sum of $\{A_\alpha\}_{\alpha\in I}$ and if

$$A_\alpha \bigcap \left(\sum_{\beta \in I \smallsetminus \{\alpha\}} A_\beta \right) = \{0\}$$

for all $\alpha \in I$.

Remark 3.25. If V is the internal direct sum of $\{A_\alpha\}_{\alpha\in I}$, then V equipped with the family $\{\iota_\alpha\}_{\alpha\in I}$ of the canonical injection maps $\iota_\alpha : A_\alpha \to V$ given by setting $\iota_\alpha(x) = x$ for all $x \in A_\alpha$ is easily seen to be a coproduct. Therefore, V is linearly isomorphic to $\bigoplus_{\alpha\in I} A_\alpha$, whence we use the same $\oplus$ notation for internal direct sums.

Note, nevertheless, that $A \oplus B$ as an internal sum *equals* V, whereas $A \oplus B$ as an external sum is merely *isomorphic* to V. Our notations in these notes shall reflect this difference.

Since finite direct sums in $F-$**Vect** agree with finite direct products, we expect there to be natural projection maps from a finite internal direct sum onto its summands. The following proposition spells this out.

Proposition 3.26. *Let V be a vector space over a field F and assume that V can be written as the internal direct sum $V = \bigoplus_{n=1}^N A_n$. There exist surjective F-linear mappings $p_n : V \to A_n$ such that*

$$p_n(x) = \begin{cases} x & \text{if } x \in A_n; \\ 0 & \text{if } x \notin A_n; \end{cases}$$

*for all n. Moreover, $(V, \{p_n\}_{n=1}^N)$ is a product of $\{A_n\}_{n=1}^N$ in $F-$**Vect**. Therefore,*

$$V \cong \prod_{n=1}^N A_n \cong \bigoplus_{n=1}^N A_n.$$

Proof. Fix $1 \leq n \leq N$. Proposition 3.9 furnishes a basis $\{v_\alpha\}_{\alpha\in I_1}$ of A_n. Another application of Proposition 3.9 furnishes a collection $\{w_\beta\}_{\beta\in I_2}$ of vectors in V such that $\{v_\alpha\}_{\alpha\in I_1} \cup \{w_\beta\}_{\beta\in I_2}$ is a basis of V. We define $p_n : V \to A_n$ by setting

- $p_n(v_\alpha) = v_\alpha$ for all $\alpha \in I_1$;
- $p_n(w_\beta) = 0$ for all $\beta \in I_2$;

and extending linearly. Evidently, $p_n|_{A_n} = \mathrm{id}_{A_n}$.

To complete the proof, we let $(Q, \{q_n\}_{n=1}^N)$ be a pair consisting of a vector space Q and a family $\{q_n\}_{n=1}^N$ of F-linear mappings $q_n : Q \to A_n$. We define $q : Q \to V$ by setting

$$q(x) = q_1(x) + \cdots + q_N(x)$$

for all $x \in Q$. Since $q_1, \ldots, q_N$ are F-linear, q is F-linear as well. Moreover,

$$p_n(q_m(x)) = \begin{cases} q_m(x) & \text{if } m = n; \\ 0 & \text{if } m \neq n; \end{cases}$$

and so

$$p_n(q(x)) = p_n(q_1(x)) + \cdots + p_n(q_N(x)) = p_n(q_n(x)) = q_n(x)$$

for all $1 \leq n \leq N$. It follows that V is a product of $\{A_n\}_{n=1}^N$ in $F-\mathbf{Vect}$.

The final string of isomorphisms now follows from the uniqueness of product (Proposition 17.15) and the uniqueness of coproduct (Proposition 17.18). □

If we think of p_n as an operator from V into itself, then $p_n \circ p_n = p_n$ for all $1 \leq n \leq N$ in the above proposition. We abstract this property to make the following definition:

Definition 3.27. Let V be a vector space over a field F. A *projection* on V is a linear operator $T : V \to V$ such that $T^2 = T$.

Proposition 3.28. *If $T : V \to V$ is a projection, then there exists a linear subspace M of T such that $\operatorname{im} T = M$ and $T|_M = \operatorname{id}_M$. Subsequently, we call T a* projection of V onto M.

Remark 3.29. If V splits as an internal direct sum $M \oplus N$, then Proposition 3.26 implies that there are projections $p_M : V \to M$ and $p_N : V \to N$.

Proof. Let $M = \operatorname{im} T$. If $x \in M$, then we can find $y \in V$ such that $Ty = x$. Since $T^2 = T$, we see that $x = Ty = TTy = Tx$. We conclude that $T|_M = \operatorname{id}_M$. □

Let us now pursue further the idea of creating a "generic" vector space of dimension κ. Given a nonempty set X of cardinality κ, we shall construct an F-vector space $\mathcal{F}_F(X)$ that has X as its natural basis. By Theorem 3.15, $\mathcal{F}_F(X)$ is necessarily isomorphic to the κ-fold direct sum $F^{\oplus X}$ of F.

Definition 3.30. Let F be a field and X a nonempty set. The *free F-vector space over X* is the set $\mathcal{F}_F(X)$ of functions $f : X \to F$ such that the cardinality of the preimage $f^{-1}(F \smallsetminus \{0\})$ is finite. Vector addition on $\mathcal{F}_F(X)$ is given by pointwise addition

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

and scalar multiplication on $\mathcal{F}_F(X)$ is given by pointwise multiplication

$$(af)(x) = af(x).$$

Proposition 3.31. *$\mathcal{F}_F(X)$ is an F-vector space of dimension $|X|$. Consequently, $\mathcal{F}_F(X) \cong F^{\oplus X}$.*

Proof. For each $x \in X$, we define the function $f_x : X \to F$ by setting

$$f_x(y) = \begin{cases} 1 & \text{if } y = x; \\ 0 & \text{if } y \neq x. \end{cases}$$

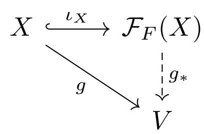
$\{f_x\}_{x \in X}$ is evidently a basis of $\mathcal{F}_F(X)$. □

Identifying f_x with x, we can think of $\mathcal{F}_F(X)$ as the space of finite "formal sums"

$$\sum_{n=1}^{N} a_n x_n$$

of arbitrary elements $x_1, \ldots, x_n$ of X. The identification process satisfies the following universal property:

Proposition 3.32. *Let F be a field and X a nonempty set. Define $\iota_X : X \to \mathcal{F}_F(X)$ by setting $\iota(x) = f_x$, where f_x is given in the proof of Proposition 3.31. Given an arbitrary function $g : X \to V$ from X to an F-vector space V, there exists a unique F-linear mapping $g_* : \mathcal{F}_F(X) \to V$ such that the diagram*



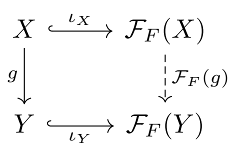
commutes.

Proof. We define $g_* : \mathcal{F}_F(X) \to V$ by setting

$$g(x) = g_*(\iota(x))$$

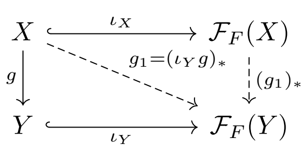
for all $x \in X$ and extending linearly. g_* is clearly the unique F-linear mapping that makes the diagram commute. □

Proposition 3.33. *Let F be a field, and let X and Y be nonempty sets. Each function $g : X \to Y$ admits a unique F-linear mapping $\mathcal{F}_F(g) : \mathcal{F}_F(X) \to \mathcal{F}_F(X)$ such that the diagram*



commutes.

Proof. We apply Proposition 3.32 twice: first to $\iota_Y \circ g$ to construct $g_1 = (\iota_Y \circ g)_*$, and then to g_1 to construct $(g_1)_*$. See the diagram below:



$\mathcal{F}_F(g) = (g_1)_*$ is the desired map. Uniqueness is guaranteed by the uniqueness clause in Proposition 3.32. □

4. Matrices

We now focus our attention to finite-dimensional vector spaces on them. All vector spaces in this section are assumed to be finite-dimensional.

Let F be a field. Recall the construction of F^n from Example 2.4 and the construction of the standard coordinate vectors $\{e_1, \ldots, e_n\}$ from Example 3.6. The *standard matrix representation* of a linear transformation $S : F^n \to F^m$ is the m-by-n array

$$\left[Se_1 \middle| \cdots \middle| Se_n \right].$$

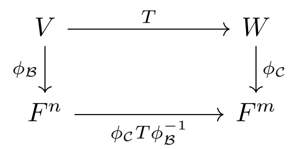
In light of this, we define an *m-by-n matrix with entries in F* to be a linear transformation from F^n to F^m and use the function notation interchangeably with the array notation presented above. We say that M is a *square matrix* if $m = n$.

Let V be a vector space over F. Given a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of a vector space V, we define the *coordinate map on V with respect to $\mathcal{B}$* to be the map $\phi_{\mathcal{B}} : V \to F^n$ given by the formula

$$\phi_{\mathcal{B}} \left(\sum_{k=1}^{n} a^k v_k \right) = \sum_{k=1}^{n} a^k e_k.$$

We write $[v]_{\mathcal{B}}$ to denote $\phi_{\mathcal{B}}(v)$.

We now let W be another vector space over F and $\mathcal{C} = \{w_1, \ldots, w_m\}$ a basis of W. For each linear transformation $T : V \to W$, the following diagram commutes:



The *matrix representation of T with respect to $\mathcal{B}$ and $\mathcal{C}$* is the standard matrix representation of $\phi_{\mathcal{C}} \circ T \circ \phi_{\mathcal{B}}^{-1}$:

$$[T]_{\mathcal{B},\mathcal{C}} = \left[[Tv_1]_{\mathcal{C}} \middle| \cdots \middle| [Tv_n]_{\mathcal{C}} \right].$$

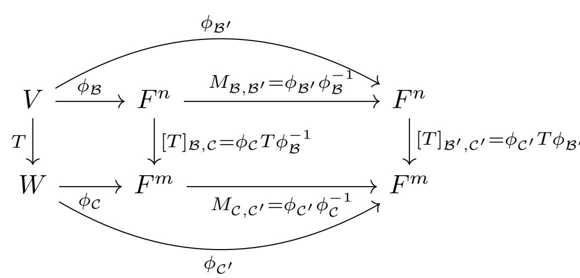
Let $\mathcal{B}' = \{v_1', \ldots, v_n'\}$ and $\mathcal{C}' = \{w_1', \ldots, w_m'\}$ be bases of V and W, respectively. The n-by-n matrix

$$M_{\mathcal{B},\mathcal{B}'} = [\mathrm{id}_{V \to V}]_{\mathcal{B},\mathcal{B}'} = \phi_{\mathcal{B}'} \circ \phi_{\mathcal{B}}^{-1}$$

is called the *change-of-basis matrix from $\mathcal{B}$ to $\mathcal{B}'$*. Indeed, we have the identity

$$[T]_{\mathcal{B}',\mathcal{C}'} = M_{\mathcal{C},\mathcal{C}'} [T]_{\mathcal{B},\mathcal{C}} M_{\mathcal{B},\mathcal{B}'}^{-1}, \tag{4.1}$$

which is derived from the commutative diagram



In particular, if $V = W$, $\mathcal{B} = \mathcal{C}$, and $\mathcal{B}' = \mathcal{C}'$, then we see that

$$[T]_{\mathcal{B}'} = M_{\mathcal{B},\mathcal{B}'} [T]_{\mathcal{B}} M_{\mathcal{B},\mathcal{B}'}^{-1}, \tag{4.2}$$

where $[T]_{\mathcal{B}} = [T]_{\mathcal{B},\mathcal{B}}$ and $[T]_{\mathcal{B}'} = [T]_{\mathcal{B}',\mathcal{B}'}$. We note that every invertible matrix can be written as a change-of-basis matrix. Indeed, if $M : F^n \to F^n$, is an invertible matrix, then M is the change-of-basis matrix from $\{e_1, \ldots, e_n\}$ to $\{Me_1, \ldots, Me_n\}$.

Two matrices A and B are *equivalent* if they represent the same linear transformation with respect to different bases. In other words, A and B are equivalent if and only if there exist invertible matrices P and Q such that

$$B = PAQ^{-1}.$$

Compare the above with (4.1).

Two square matrices A and B are *similar* if there exists an invertible matrix P such that

$$B = PAP^{-1}.$$

Compare the above with (4.2). Analogously, we say that two linear operators $C, D : V \to V$ are *similar* if there exists an invertible linear transformation $\phi : V \to V$ such that

$$D = \phi C \phi^{-1}.$$

Two matrices are similar if and only if they are matrix representations of the same linear operator, and two linear operators are similar if and only if they are represented by the same matrix.

We now introduce basic operations on matrices. Given an m-by-n matrix

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \ddots & & & \vdots \\ \vdots & & \ddots & & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix},$$

we write M_{ij} to indicate a_{ij}. Given a collection $\{a_{ij} : 1 \leq i \leq m \text{ and } 1 \leq j \leq n\}$ of elements in F, we write $(a_i j)$ to represent the m-by-n matrix M such that $M_{ij} = a_{ij}$ for all choices of i and j.

The *sum* of two m-by-n matrices $A = (a_{ij})$ and $B = (b_{ij})$ is the sum of A and B as linear operators in $\mathscr{L}(F^n, F^m)$; see Definition 6.1. Likewise, the *product* of a scalar $\lambda \in F$ and A is the product of λ and A in $\mathscr{L}(F^n, F^m)$. If C is an n-by-p matrix, then the *product* of A and C is the function composition $A \circ C \in \mathscr{L}(F^p, F^m)$ of $A \in \mathscr{L}(F^n, F^m)$ and $C \in \mathscr{L}(F^n, F^p)$. It is not difficult to check that

$$A + B = (a_{ij} + b_{ij}), \qquad \lambda A = (\lambda a_{ij}), \qquad \text{and} \qquad AC = \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right).$$

Definition 4.3. The *transpose* of an m-by-n matrix $M = (a_{ij})$ is the n-by-m matrix $M^t = (a_{ji})$. M is said to be a *symmetric matrix* if $m = n$ and $M^t = M$.

Proposition 4.4. *Let A and B be m-by-n matrices with entries in a field F, let C be an n-by-p matrix with entries in F, and $\lambda \in F$.*

(1) $(A^t)^t = A$.
(2) $(A + B)^t = A^t + B^t$.
(3) $(\lambda A)^t = \lambda A^t$.
(4) $(AC)^t = C^t A^t$.
(5) $A^t = A^\star$, *where $A^\star$ is the algebraic adjoint (Definition 6.11) of A as a linear operator.*
(6) If A is an invertible square matrix, then $(A^t)^{-1} = (A^{-1})^t$.

Proof. (1) - (4) are trivial. For (5), we observe that

$$A^t (A^{-1})^t = (A^{-1} A)^t = I^t = I.$$

Similarly, $(A^{-1})^t A^t = I$. ∎

5. Quotient Spaces

Let us now study the basic properties of very general linear operators, without any additional assumptions.

We begin by introducing two fundamental subspaces associated with T.

Definition 5.1. Let $T : V \to W$ be a linear operator between two vector spaces V and W over a field F. The *range*, or the *image* of T is the set

$$\operatorname{im} T = \{Tv : v \in V\},$$

and the *nullspace*, or the *kernel*, of T is the set

$$\ker T = \{v \in V : Tv = 0\}.$$

Remark 5.2. It is easy to see that $\operatorname{im} T$ is a linear subspace of W, and that $\ker T$ is a linear subspace of V.

Every linear operator $T : V \to W$ can be turned into a surjective one by restricting the codomain: $T : V \to \operatorname{im} T$. Note now that $Tv = Tw$ implies that $v - w \in \ker T$. Therefore, if we identify all elements of $\ker T$, then we get the "injective version" of T, thus obtaining an isomorphism. Let us now make this precise.

Definition 5.3. Let V be a vector space over a field F, and W a subspace of V. The *quotient space* V/W is defined to be the quotient set $V/\sim$ with respect to the equivalence relation

$$v \sim w \quad \text{if and only if} \quad v - w \in W,$$

endowed with vector addition

$$[v] + [w] = [v + w]$$

and scalar multiplication

$$a[v] = [av].$$

The *canonical surjection* $p : V \to V/W$ is defined by setting

$$p(v) = [v]$$

for each $v \in V$.

Remark 5.4. p is easily seen to be a surjective linear operator.

Remark 5.5. Note that the equivalence class $[v] \in V/W$ is precisely the *coset*

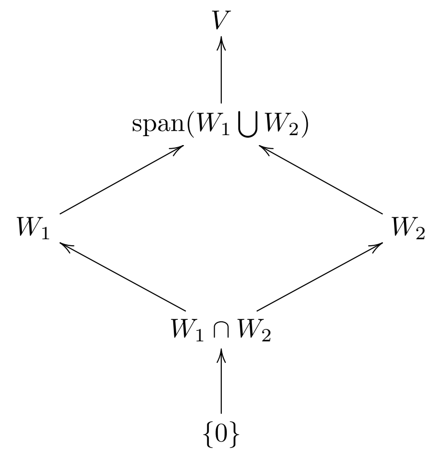
$$\{v + w : w \in W\}.$$

We typically write $v + W$ to denote the above coset.

We observe that quotienting V out by W collapses every element of W onto the zero vector. Indeed, the equivalence class $[0]$ consists precisely of the vectors in W. It is then natural to expect that the quotient space V/W is smaller if W is larger. More precisely, if we define a partial order on the set of all linear subspaces of V by setting

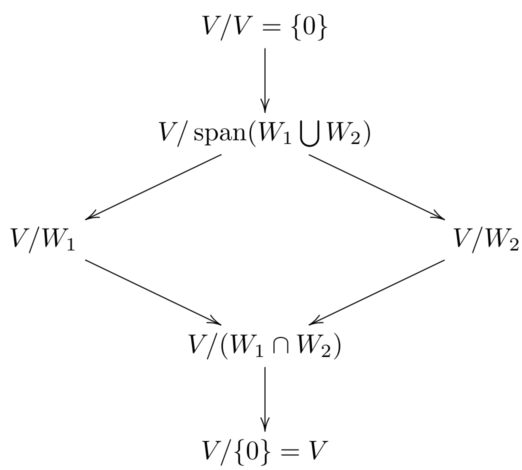
$$W_1 \leq W_2 \text{ if and only if } W_1 \text{ is a linear subspace of } W_2, \tag{5.6}$$

then, given any two linear subspaces W_1 and W_2 of V, we can draw a *lattice of linear subspaces*



where the arrows indicate an increase in size with respect to the partial order $\leq$.

We contend that the act of quotienting out reverses the arrows in the lattice, i.e.,



where the arrows now indicate the "embeds to" relation. The precise statement is as follows:

Proposition 5.7. *The relation*

$$V/W_1 \lesssim V/W_2 \Leftrightarrow \text{ there exists an embedding } V/W_1 \to V/W_2 \tag{5.8}$$

is a reflexive, transitive relation on the set quotient spaces of V such that $W_1 \geq W_2$ in the sense of (5.6) implies $V/W_1 \lesssim V/W_2$. □

We omit the proof of the above proposition. Note that $\lesssim$ is not quite a partial order, for it is possible for two non-equal quotient spaces to be isomorphic, thereby admitting embeddings in both directions. The correct set-up would be to consider *isomorphism classes* of subspaces of V and quotient spaces of V, respectively, and consider the partial orders on these sets given by the embedding relation (5.8). The antisymmetry property that failed in the setting of Proposition 5.7 now holds, and this is a consequence of the following theorem:

Theorem 5.9 (Schröder-Bernstein for vector spaces)**.** *If two vector spaces X and Y admit embeddings $X \to Y$ and $Y \to X$, then X is linearly isomorphic to Y.*

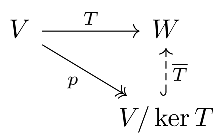
See [MO1] for a discussion on the Schröder-Bernstein property on different categories. We do not pursue this matter further in these notes.

Returning to the topic at hand, we observe that shrinking the vector space reduces the size of the corresponding quotient space. Once again, we omit the proof of the following proposition.

Proposition 5.10. *Let V be a vector space and W a linear subspace of V. Every linear subspace of the quotient space V/W is of the form X/W, where $W \leq X \leq V$ in the sense of (5.6). Conversely, if X is a linear subspace of V containing W, then X/W is a linear subspace of V/W.* □

With this we now leave behind the order properties of quotient spaces and establish the basic tools of the trade.

Theorem 5.11 (First isomorphism theorem for vector spaces)**.** *If $T : V \to W$ is a linear operator, then there exists a unique injective linear operator $\overline{T} : V/\ker T \to W$ such that the diagram*



commutes. In particular, $V/\ker T$ is linearly isomorphic to $\operatorname{im} T$.

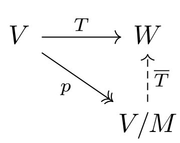
Proof. Define $\overline{T} : V/\ker T \to W$ by setting $\overline{T}(v + W) = Tv$ for each $v \in V$. If $v_1, v_2 \in (v + W)$, then $v_1 - v, v_2 - v \in \ker T$, and so

$$Tv_1 - Tv_2 = T(v_1 - v + v - v_w) = T(v_1 - v_2) = 0.$$

Therefore, $\overline{T}$ is well-defined. $\overline{T}$ is easily seen to be an injective linear operator. The uniqueness clause follows at once from the definition. □

The same method of proof yields more, in fact:

Theorem 5.12 (Universal property of quotient vector spaces)**.** *Let V and W be vector spaces over a field F and M a linear subspace of V. If $T : V \to W$ is a linear operator such that $M \subseteq \ker T$, then there exists a unique linear operator $\overline{T} : V/M \to W$ such that the diagram*



commutes. Moreover, $\ker \overline{T} = \ker T/M$ and $\operatorname{im} \overline{T} = \operatorname{im} T$.

Proof. Define $\overline{T} : V/M \to W$ by setting $\overline{T}(v + M) = Tv$ for each $v \in V$. Since $M \subseteq \ker T$, we repeat the proof of the first isomorphism theorem (Theorem 5.11) to conclude that $\overline{T}$ is well-defined and unique. The identities $\ker \overline{T} = \ker T/M$ and $\operatorname{im} \overline{T} = \operatorname{im} T$ follow at once from the construction. □

We now apply the first isomorphism theorem to the study of complements of subspaces, which we now define.

Definition 5.13. Let V be a vector space over a field F, and M a linear subspace of V. A *complement of M in V* is a linear subspace N of V such that

$$V = M \oplus N.$$

Remark 5.14. It is easy to see that complements exist: just take a basis of M and extend it to a basis of V via Theorem 3.9. This is not necessarily true for *Banach spaces* or *Hilbert spaces*.

An immediate consequence, by way of Remark 3.29, is that, if V is a vector space and M is a linear subspace of V, then there exists a projection of V onto M.

Corollary 5.15 ("Splitting lemma" for vector spaces)**.** *Let V be a vector space over a field F, and M a linear subspace of V. Every complement of M is isomorphic to V/M. Therefore,*

$$V \cong M \oplus V/M.$$

Proof. Suppose that N is a complement of M in V. We know from Proposition 3.26 that there exists a surjective F-linear mapping $p : V \to N$ such that $\ker p = M$. The first isomorphism theorem (Theorem 5.11) implies that

$$V/M = V/\ker p \cong \operatorname{im} p = N,$$

as was to be shown.

Let $\overline{p} : V/M \to N$ be the linear isomorphism. The mapping $T : M \oplus V/M \to V$ given by setting

$$T(v, w) = v + \overline{q}(w)$$

is easily seen to be an isomorphism. □

Remark 5.16. An equivalent formulation of the above corollary is that

$$A \oplus B = C \oplus D$$

with $A = C$ implies that $B \cong D$.

Example 5.17. Note, however, that

$$A \oplus B = C \oplus D$$

with $A \cong C$ does *not* imply that $B \cong D$. To see this, we shall show that

$$F^{\oplus \aleph_0} \cong F^{\oplus \aleph_0} \oplus F^{\oplus \aleph_0} \tag{5.18}$$

for any choice of field F, where $\aleph_0 = \mathbb{N}$ is the first infinite cardinal. Since

$$F^{\oplus \aleph_0} \oplus F^{\oplus \aleph_0} = \left(F^{\oplus \aleph_0} \oplus F^{\oplus \aleph_0}\right) \oplus \{0\},$$

the above claim furnishes a requisite counterexample.

Let O be the set of odd numbers in $\mathbb{N}$, and let E be the set of even numbers in $\mathbb{N}$. Evidently,

$$F^{\oplus \aleph_0} = F^{\oplus O} \oplus F^{\oplus E}.$$

But $|O| = \aleph_0$ and $|E| = \aleph_0$, and so Theorem 3.15 implies that $F^{\oplus O} \cong F^{\oplus \aleph_0}$ and $F^{\oplus E} \cong F^{\oplus \aleph_0}$. It follows that (5.18) holds. □

Example 5.19. (5.18) also shows that

$$A \oplus B \cong C \oplus D$$

with $A = C$ does not imply that $B \cong D$. This, in conjunction with Corollary 5.15, implies that $V \cong W$ does not necessarily imply $V/M \cong W/M$, where M is a shared linear subspace of V and W. □

Yet another way of writing the splitting lemma (Corollary 5.15) is that

$$(A \oplus B)/B \cong A/(A \cap B),$$

as $A/A \cap B = A/\{0\} = A$ in the internal direct sum case. We now show that the same statement holds for Minkowski sums:

Corollary 5.20 (Second isomorphism theorem for vector spaces)**.** *If V is a vector space over F, and if M and N are linear subspaces of V, then*

$$(M + N)/N \cong M/(M \cap N).$$

Proof. We define $T : M + N \to M/(M \cap N)$ by setting

$$T(v + w) = v + (M \cap N)$$

for each $v \in M$ and every $w \in N$. T is easily seen to be a surjective linear mapping. Since $\ker T = N$, the first isomorphism theorem (Theorem 5.11) implies that $(M + N)/N \cong M/(M \cap N)$. □

To what extent do quotient spaces behave like fractions? The following corollaries of the first isomorphism theorem shed a light on this matter.

Corollary 5.21 (Third isomorphism theorem for vector spaces)**.** *Let V be a vector space over a field F. If M and N are linear subspaces of V such that $N \subseteq M \subseteq V$, then*

$$(V/N)/(M/N) \cong V/M.$$

Proof. Apply the first isomorphism theorem to the mapping $T : V/N \to V/M$ defined by setting $T(v + N) = v + M$. □

Corollary 5.22. *Let V_1 and V_2 be vector spaces over a field F, and let M_1 and M_2 be linear subspaces of V_1 and V_2, respectively. Then*

$$(V_1 \oplus V_2)/(M_1 \oplus M_2) \cong (V_1/M_1) \oplus (V_2/M_2).$$

Proof. We apply the first isomorphism theorem to the mapping $T : V_1 \oplus V_2 \to (V_1/M_1) \oplus (V_2/M_2)$ defined by setting $T(v_1 + v_2) = (v_1 + M_1, v_2 + M_2)$. □

6. Duality

We now turn to the study of $\mathrm{Hom}_{F-\mathbf{Vect}}(V, W)$, the collection of all linear operators from an F-vector space V to another F-vector space W. Of particular interest is $\mathrm{Hom}_{F-\mathbf{Vect}}(V, F)$, which we call the *dual space* of V.

Definition 6.1. Let V and W be vector spaces over a field F. $\mathscr{L}_F(V, W)$ denotes $\mathrm{Hom}_{F-\mathbf{Vect}}(V, W)$, the set of all linear operators from V to W. We often omit F and write $\mathscr{L}(V, W)$.

Remark 6.2. We can turn $\mathscr{L}(V, W)$ into an F-vector space by defining

$$(T + S)x = Tx + Sx \quad \text{and} \quad (aT)x = a(Tx)$$

for all $T, S \in \mathscr{L}(V, W)$, $x \in V$, and $a \in F$.

Theorem 6.3. $\dim \mathscr{L}(V, W) \geq \dim V \cdot \dim W$. *The equality holds if and only if* $\dim V < \infty$ *or* $\dim W = 0$.

Proof. If $\dim W = 0$, then the theorem holds trivially. We therefore assume that $\dim W > 0$.

Let $\{v^\alpha\}_{\alpha\in I}$ and $\{w^\beta\}_{\beta\in J}$ be bases of V and W, respectively. We define $\{T^{\alpha\beta}\}_{\alpha\in I,\beta\in J} \subseteq \mathscr{L}(V,W)$ by setting

$$T^{\alpha\beta}v = \begin{cases} aw^\beta & \text{if } v = av^\alpha \text{ for some } a \in F; \\ 0 & \text{otherwise;} \end{cases}$$

which is a linearly independent subset of $\mathscr{L}(V,W)$. Therefore, $\dim \mathscr{L}(V,W) \geq \dim V \cdot \dim W$.

If V is infinite-dimensional, then we can find a partition $\{I_1, I_2\}$ of I such that both I_1 and I_2 are infinite sets. We fix $\beta_0 \in J$ and define a linear operator $T : V \to W$ by setting

$$Tv^\alpha = \begin{cases} w^{\beta_0} & \text{if } \alpha \in I_1; \\ 0 & \text{if } \alpha \in I_2 \end{cases}$$

and extending linearly. It is not hard to see that T fails to be in the span of $\{T^{\alpha\beta}\}_{\alpha\in I,\beta\in J}$, whence $\dim \mathscr{L}(V,W) > \dim V \cdot \dim W$.

If V is finite-dimensional, then we can take I to be $\{1,\ldots,N\}$ for some $N \in \mathbb{N}$. Given $T \in \mathscr{L}(V,W)$, we take the basis expansion

$$Tv^n = \sum_{\beta\in J} a_{n,\beta} w^\beta$$

for each $1 \leq n \leq N$. The sum

$$\sum_{n=1}^{N} \sum_{\beta\in J} a_{n,\beta} T^{n,\beta}$$

is finite and equals T. It follows that $\dim \mathscr{L}(V,W) = \dim V \cdot \dim W$. □

Definition 6.4. The *(algebraic) dual space* $V^\star$ of a vector space V over a field F is the space $\mathscr{L}(V,F)$. An element of $\mathscr{L}(V,F)$ is referred to as a *linear functional on* V.

Example 6.5. Let V be an F-vector space of F-valued functions on a set X. Given $x \in X$, the *evaluation functional at* X, defined for each $f \in V$ by the formula

$$l_x(f) = f(x),$$

is a linear functional on V. □

The operators $T^{\alpha\beta}$ constructed in the proof of Theorem 6.3 gives rise to the standard basis on $V^\star$.

Corollary 6.6. *Given a basis $\{v_\alpha\}_{\alpha\in I}$ of V, the set $\{v_\alpha^\star\}_{\alpha\in I}$ of linear functionals on V given by the linear extension of the formula*

$$v_\alpha^\star(v^\beta) = \begin{cases} 1 & \text{if } \alpha = \beta; \\ 0 & \text{if } \alpha \neq \beta; \end{cases}$$

is a linearly independent subset of $V^\star$, called the dual basis *of $V^\star$ with respect to $\{v_\alpha\}_{\alpha\in I}$. The dual basis is a basis of $V^\star$ if and only if V is finite-dimensional. Indeed,* $\dim V^\star \geq \dim V$ *and the equality holds if and only if* $\dim V < \infty$.

Proof. This is a direct consequence of Theorem 6.3. Alternatively, we observe that $V \cong F^{\oplus I}$ and $V^\star \cong F^I$, whence the desired result follows from Theorem 3.20. $\square$

The construction of the dual basis is not *natural*, in the sense that it depends on the choice of a basis. Compare with the following result for the double dual. (See §17.3 for category-theoretical details.)

Definition 6.7. The *(algebraic) double dual* $V^{\star\star}$ of a vector space V of a field F is the space $\mathscr{L}(V^\star, F)$.

Theorem 6.8. *Given an element $x \in V$ of the vector space V over a field F, the* evaluation functional $\hat{x} : V^\star \to F$ *defined by setting*

$$\hat{x}(l) = l(x)$$

is an element of $V^{\star\star}$. The mapping $x \mapsto \hat{x}$ is an injective linear operator from V into $V^{\star\star}$ and is a linear isomorphism if and only if V is finite-dimensional.

To prove the theorem, we shall make use of an important lemma. While the lemma might seem trivial in the context of linear algebra, While it might seem trivial in this context, the lemma has a functional-analytic variant (Theorem 11.12) that serves as one of the cornerstones of Banach space theory.

Lemma 6.9 (Algebraic Hahn-Banach theorem)**.** *For each nonzero vector v in a vector space V, there exists a linear functional $l_v \in V^\star$ such that $l_v(v) = 1$ and that $l_v(w) = 0$ for all $w \in V \smallsetminus \operatorname{span}\{v\}$.*

Proof of lemma. We find, via Theorem 3.9, a basis $\{v_\alpha\}_{\alpha \in I}$ of V such that $v = v_{\alpha_0}$ for some fixed $\alpha_0 \in I$. The linear functional defined by setting

$$l_v\left(\sum_{\alpha \in I} a_\alpha v_\alpha\right) = a_{\alpha_0}$$

is the desired linear functional. $\square$

Proof of theorem. Suppose that $\widehat{x_1} = \widehat{x_2}$, so that $l(x_1 - x_2) = 0$ for all $l \in V^\star$. If $x_1 - x_2 \neq 0$, then the algebraic Hahn-Banach theorem furnishes an $l_0 \in V^\star$ such that $l_0(x_1 - x_2) \neq 0$, which is absurd. It follows that $x_1 = x_2$, whence the mapping $x \mapsto \widehat{x}$ is injective.

If V is finite-dimensional, then it follows from Corollary 6.6 that $V \cong V^\star \cong V^{\star\star}$. It now follows from the splitting lemma (Corollary 5.15) that the mapping $x \mapsto \widehat{x}$ must be surjective as well. $\square$

Remark 6.10. Even though the proof of the algebraic Hahn–Banach theorem depends on the *existence* of a basis, we stress that the construction of the evaluation functionals does not depend on the *choice* of a basis.

The construction of evaluation functionals given above is typical. Many important new functions are built out of old functions via analogous constructions. For example, the same construction gives the *dual* operator $T^\star : W^\star \to V^\star$ of a linear operator $T : V \to W$, which we now define.

Definition 6.11. Let V and W be vector spaces over a field F. The *transpose* of $T \in \mathscr{L}(V, W)$ is the linear operator $T^\star \in \mathscr{L}(W^\star, V^\star)$ defined by the formula

$$T^\star(l) = l \circ T.$$

$T^\star$ is also referred to as the *dual operator of T* or the *algebraic adjoint of T*.

Here is a simple but important example of a duality property.

Theorem 6.12. *Let V and W be vector spaces over a field F, and let $T \in \mathscr{L}(V, W)$. If T is injective, then $T^\star$ is surjective. If T is surjective, then $T^\star$ is injective.*

For the proof, we shall need the algebraic Hahn-Banach theorem, as well as its partial converse.

Lemma 6.13. *For each nonzero linear functional $l \in V^\star$, there exists a nonzero vector $v \in V$ such that $l(v) = 1$ and that $l(w) = 0$ for all $w \in V \smallsetminus \operatorname{span}\{v\}$.*

Proof of lemma. Since l is nonzero, l is a surjective linear operator from V to the ground field F. The first isomorphism theorem (Theorem 5.11) implies that $V/\ker l \cong F$. As per the splitting lemma (Corollary 5.15), every complement of $\ker l$ is isomorphic to $V/\ker l$ and is therefore one-dimensional. The desired result now follows. □

The proof of the theorem is now straightforward.

Proof of theorem. Suppose that T is injective. Given $l \in V^\star$, Lemma 6.13 furnishes a vector $v \in V^\star$ such that $l(v) = 1$ and that $l(w) = 0$ for all $w \in V \smallsetminus \operatorname{span}\{v\}$. Since T is injective, Tv is nonzero, whence the algebraic Hahn-Banach theorem (Lemma 6.9) furnishes a linear functional $L_l \in W^\star$ such that $L_l(Tv) = 1$ and that $L_l(w) = 0$ for all $w \in W \smallsetminus \operatorname{span}\{Tv\}$. By construction, $T^\star(L_l) = L_l \circ T = l$. Since l was arbitrary, we conclude that $T^\star$ is surjective.

We now suppose instead that T is surjective. Suppose that $l_1, l_2 \in W^\star$ satisfy the identity $T^\star l_1 = T^\star l_2$. By definition, $l_1 \circ T = l_2 \circ T$, and so $l_1|_{\operatorname{im} T} = l_2|_{\operatorname{im} T}$. Since $\operatorname{im} T = W$, it follows that $l_1 = l_2$, whence $T^\star$ is injective. □

As an application of the theorem, we study the dual of quotient spaces. Given a vector space V and a linear subspace M of V, we have the canonical surjection $p : V \to V/M$. The injection-surjection duality implies that the transpose $p^\star : (V/M)^\star \to V^\star$ is injective, whence $(V/M)^\star$ is linearly isomorphic to a subspace of $V^\star$. By definition, each $l \in (V/M)^\star$, satisfies the identity $p^\star l = l \circ p$, whence $p^\star l$ must vanish on $\ker p = M$. We thus make the following definition:

Definition 6.14. Let M be a subset of a vector space V. The *annihilator of M* is the set

$$M^\circ = \{l \in V^\star : l|_M = 0\}.$$

Dualizing the canonical maps, we obtain the following isomorphism theorems:

Theorem 6.15 (Isomorphism theorems for annihilators)**.** *Let M be a linear subspace of a vector space V.*

(1) $(V/M)^\star \cong M^\circ$.
(2) $V^\star/M^\circ \cong M^\star$.

Proof. (1) Consider the canonical surjection $p : V \to V/M$. By the injection-surjection duality (Theorem 6.12) the transpose $p^\star : (V/M)^\star \to V^\star$ is injective. It now suffices to note that $\operatorname{im} p^\star = M^\circ$.

(2) Consider the canonical injection map $\iota : M \to V$. By the injection-surjection duality (Theorem 6.12), the transpose $\iota^\star : V^\star \to M^\star$ is surjective. Since $\ker \iota^\star = M^\circ$, the desired result follows from the first isomorphism theorem (Theorem 5.11). □

With annihilators, we can refine the injection-surjection duality:

Theorem 6.16 (Injection-surjection duality of the algebraic adjoint)**.** *If* $T \in \mathscr{L}(V, W)$*, then*

(1) $\ker T^\star = (\operatorname{im} T)^\circ$*, and*
(2) $\operatorname{im} T^\star = (\ker T)^\circ$*.*

Proof. (1) Observe that

$$\begin{aligned}\ker T^\star &= \{l \in W^\star : T^\star l = 0\} = \{l \in W^\star : \operatorname{im} lT = \{0\}\} \\ &= \{l \in W^\star : lT(V) = \{0\}\} = \{l \in W^\star : l(\operatorname{im} T) = \{0\}\} = (\operatorname{im} T)^\circ.\end{aligned}$$

(2) If $l \in \operatorname{im} T^\star$, then there exists $L \in W^\star$ such that $LT = l$. For each $v \in \ker T$, we have the identity $lv = LTv = L(0) = 0$. It follows that $l \in (\ker T)^\circ$. Since l was arbitrary, we conclude that $\operatorname{im} T^\star \subseteq (\ker T)^\circ$.

To show the reverse inclusion, we fix $l \in (\ker T)^\circ$. By Lemma 6.13, there exists a nonzero vector $v \in V$ such that $l(v) = 1$ and that $l(w) = 0$ for all $w \in V \smallsetminus \operatorname{span}\{v\}$. Since $\ker l \supseteq \ker T$, we see that $Tv \neq 0$. The algebraic Hahn-Banach theorem (Lemma 6.9) now furnishes a nonzero linear functional $L \in W^\star$ such that $L(Tv) = 1$ and that $L(w) = 0$ for all $w \in W \smallsetminus \operatorname{span}\{Tv\}$. By construction, $T^\star(L) = LT = l$, whence $l \in \operatorname{im} T^\star$. Since l was arbitrary, we conclude that $(\ker T)^\circ \subseteq \operatorname{im} T^\star$. □

We conclude this section by studying the dual of direct sums.

Theorem 6.17. *If*

$$V = \bigoplus_{n=1}^{N} V_n,$$

then

$$V^\star = \bigoplus_{n=1}^{N} V_n^\circ.$$

Proof. By Proposition 3.26, we can find surjective F-linear mappings $p_n : V \to V_n$ such that

$$p_n(v) = \begin{cases} v & \text{if } v \in V_n; \\ 0 & \text{if } v \notin V_n. \end{cases}$$

Since $\operatorname{id}_V = p_1 + \cdots + p_N$, we see for each $l \in V^\star$ that

$$l = l \circ (p_1 + \cdots + p_N) = lp_2 + lp_3 + \cdots + lp_N + lp_1.$$

As $lp_n \in V_m^\circ$ whenever $n \neq m$, we conclude that $l \in \bigoplus_{n=1}^{N} V_n^\circ$. Therefore,

$$V^\star \subseteq \bigoplus_{n=1}^{N} V_n^\circ.$$

The reverse inclusion is trivial, as $V_1, \ldots, V_N$ are linear subspaces of $V^\star$. □

7. Tensor Products; Change of Base Field

We now take a second look at the evaluation functionals defined in Theorem 6.8. In the context of the theorem, we focused on constructing a mapping from V to $V^{\star\star}$ by defining

$$\hat{x}(l) = l(x)$$

for each $x \in V$ and every $l \in V^\star$. Since $l : V \to F$ and $\hat{x} : V^{\star\star} \to F$ are both linear, we see that the evaluation map $B : V \times V^\star \to F$ given by the formula

$$B(x, l) = l(x)$$

is linear in each variable. This observation leads us to introduce the following notion.

Definition 7.1 (Multilinear operators and forms)**.** Let $V_1, \ldots, V_n, W$ be vector spaces over a field F. An *n-linear operator on* $\prod_{k=1}^n V_k$ is a mapping $\eta : \prod_{k=1}^n V_k \to W$ that is linear in each variable, i.e., the map

$$v \mapsto \eta(v_1, \ldots, v_{k-1}, v, v_{k+1}, \ldots, v_n)$$

is a linear operator from V_k to W whenever $v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n$ are fixed. If, in addition, $W = F$, then we say that η is an *n-linear functional.* Furthermore, if $W = F$ and $V_1 = \cdots = V_n$, then we say that η is an *n-linear form*, or an *n-tensor*. If $n = 2$, then we use the adjective *bilinear* instead of *2-linear*; similarly, if $n = 3$, then we use *trilinear*.

Example 7.2. As discussed above, the evaluation map $B : V \times V^\star \to F$ given by $B(x, l) = l(x)$ is bilinear. □

Example 7.3. Let us think of $\mathbb{C}$ as a 2-dimensional vector space over $\mathbb{R}$. The *complex multiplication* operation

$$(a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i$$

is bilinear. □

Example 7.4. We let $\mathbb{C}[x]$ denote the set of all complex polynomials. With usual addition and scalar multiplication, $\mathbb{C}[x]$ is a vector space over $\mathbb{C}$. The usual multiplication of polynomials is bilinear.

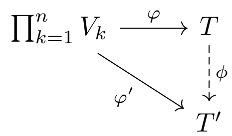
In general, polynomial multiplication on the space $F[x]$ of polynomials with coefficients in a field F is bilinear. □

Our immediate goal is to devise a canonical way of relating multilinear operators to linear operators, so as to be able to study multilinear operators with the tools of linear algebra.

Theorem 7.5 (Tensor products)**.** *Given vector spaces $V_1, \ldots, V_n$ over a field F, there exists a pair (T, φ) consisting of a vector space T over F and an n-linear mapping $\varphi : \prod_{k=1}^n V_k \to T$ such that every n-linear mapping $\eta : \prod_{k=1}^n V_k \to W$ into an arbitrary vector space W over F induces a unique linear operator $\tilde{\eta} : T \to W$, called the* mediating morphism *of η, such that the following diagram commutes:*

$$\begin{array}{ccc} \prod_{k=1}^n V_k & \xrightarrow{\varphi} & T \\ & \searrow^{\eta} & \downarrow{\scriptstyle \tilde{\eta}} \\ & & W \end{array}$$

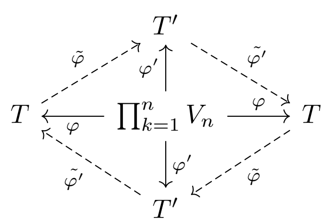
Furthermore, if (T', φ') *is another pair that satisfies the above universal property, then there exists a unique linear isomorphism* $\phi : T \to T'$ *such that the following diagram commutes:*



We write $V_1 \otimes \cdots \otimes V_n$ *or* $\bigotimes_{k=1}^n V_k$ *to denote any such vector space* T *and say that* $\bigotimes_{k=1}^n V_k$ *is the* tensor product *of the vector spaces* $V_1, \ldots, V_n$*. Given* $v_k \in V_k$ *for all* $1 \leq k \leq n$*, we write* $v_1 \otimes \cdots \otimes v_n$ *to denote the element* $\varphi(v_1, \ldots, v_n)$ *of* $\bigotimes_{k=1}^n V_k$*. If* $V = V_1 = \cdots = V_n$*, then we also write* $V^{\otimes n}$ *for* $\bigotimes_{k=1}^n V$*.*

In other words, every multilinear operator η induces a linear operator $\tilde{\eta}$ on the tensor product in a natural manner.

Proof. We suppose for now that there are two pairs (T, φ) and (T', φ') that satisfy the universal property described above. We can then find linear operators $\tilde{\varphi}$ and $\tilde{\varphi}'$ such that the following diagram commutes:



By the uniqueness clause, $\tilde{\varphi}'\tilde{\varphi} = \mathrm{id}_T$ and $\tilde{\varphi}\tilde{\varphi}' = \mathrm{id}_{T'}$. It follows that $\phi = \tilde{\varphi}$ is a linear isomorphism from T to T'. The uniqueness of ϕ also follows from the uniqueness clause of the universal property.

It remains to construct a tensor product. For notational convenience, we assume that $n = 2$. The construction for the $n > 2$ case is entirely analogous. We shall construct a tensor product of V_1 and V_2 by singling out the ordered pairs (v_1, v_2) that satisfy the desired bilinearity properties. Specifically, we start with the free vector space $\mathcal{F}_F(V_1 \times V_2)$, which was constructed in Definition 3.30. We let M be the subspace of $\mathcal{F}_F(V_1 \times V_2)$ spanned by the following vectors:

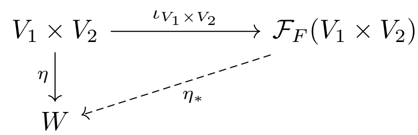
(i) $(v_1 + v_2, w) - (v_1, w) - (v_2, w)$, where $v_1, v_2 \in V_1$ and $w \in V_2$;
(ii) $(v, w_1 + w_2) - (v, w_1) - (v, w_2)$, where $v \in V_1$ and $w_1, w_2 \in V_2$;
(iii) $(av, w) - a(v, w)$, where $a \in F$, $v \in V_1$, and $w \in V_2$;
(iv) $(v, aw) - a(v, w)$, where $a \in F$, $v \in V_1$, and $w \in V_2$.

We let T be the quotient space $\mathcal{F}_F(V_1 \times V_2)/M$ and denote by $v \otimes w$ the image of $(v, w) \in \mathcal{F}(V_1 \times V_2)$ under the canonical surjection $p : \mathcal{F}_F(V_1 \times V_2) \to T$. By construction, the following properties are satisfied:

(i) $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$, where $v_1, v_2 \in V_1$ and $w \in V_2$;
(ii) $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$, where $v \in V_1$ and $w_1, w_2 \in V_2$;
(iii) $av \otimes w = a(v \otimes w)$, where $a \in F$, $v \in V_1$, and $w \in V_2$;
(iv) $v \otimes aw = a(v \otimes w)$, where $a \in F$, $v \in V_1$, and $w \in V_2$.

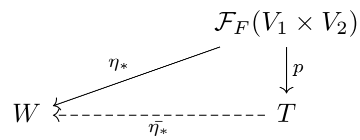
As such, the restriction $\varphi = p|_{V_1 \times V_2}$ is evidently bilinear on $V_1 \times V_2$.

Now, we consider another bilinear operator $\eta : V_1 \times V_2 \to W$ into some vector space W over F. By the universal property of free vector spaces (Proposition 3.32), there exists a unique F-linear mapping η_* that makes the following diagram commute:

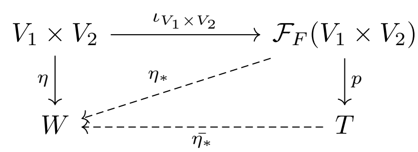


Here $\iota_{V_1 \times V_2}$ is the canonical injection map.

Now, the universal property of quotient vector spaces (Theorem 5.12) furnishes a unique linear mapping $\bar{\eta}_*$ such that the following diagram commutes:



Combining the two diagrams, we obtain the following commutative diagram:



Setting $\tilde{\eta} = \bar{\eta}_*$, we see that the universal property is satisfied. The uniqueness clause follows from the uniqueness clauses of the two universal properties we have invoked. □

For notational convenience, we shall work with the tensor product of two vector spaces for the remainder of this section unless more spaces are necessary.

Proposition 7.6. *Let V and W be vector spaces over a field F and let $(V \otimes W, \varphi)$ be the tensor product of V and W. If $\{v_\alpha\}_{\alpha \in I}$ and $\{w_\beta\}_{\beta \in J}$ are bases of V and W, respectively, then $\{v_\alpha \otimes w_\beta\}_{\alpha \in I, \beta \in J}$ is a basis of $V \otimes W$.*

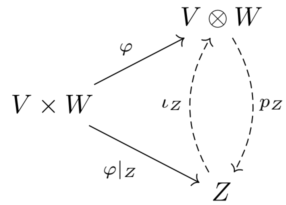
Proof. That $\{v_\alpha \otimes w_\beta\}_{\alpha \in I, \beta \in J}$ is a linearly independent subset of $V \otimes W$ is an immediate consequence of bilinearity.

Define $Z = \operatorname{span}\{v_\alpha \otimes w_\beta\}_{\alpha \in I, \beta \in J}$, which is a linear subspace of $V \otimes W$. Let X be a vector space over F and $\eta : V \times W \to X$ be a bilinear map. The map $\bar{\eta} : Z \to X$ defined by setting $\bar{\eta}(v \times w) = \eta(v, w)$ and extending linearly is a linear operator such that $\eta = \bar{\eta} \circ (\varphi|_Z)$. If $f : Z \to X$ is an arbitrary linear operator such that $\eta = f \circ (\varphi|_Z)$, then we must have

$$\eta(v, w) = (f \circ (\varphi|_Z))\,(v, w) = f(v \otimes w),$$

for all $(v, w) \in V \times W$, and so $f = \bar{\eta}$. It follows that $(Z, \varphi|_Z)$ is a tensor product of V and W.

Let $\iota_Z : Z \to V \otimes W$ be the canonical injection map, and let $p_Z : V \otimes W \to Z$ be a projection of V onto Z (Remark 5.14). It follows at once that the following diagram commutes:



Since $(V \otimes W, \varphi)$ and $(Z, \varphi|_Z)$ are tensor products of V and W, the uniqueness clause in Theorem 7.5) implies that $p_Z \circ \iota_Z = \mathrm{id}_{V \otimes W}$ and $\iota_Z \circ p_Z = \mathrm{id}_Z$. This is possible only if $Z = V \otimes W$. □

It now follows that every $x \in V \otimes W$ has a representation of the form $v_1 \otimes w_1 + \cdots + v_n \otimes w_n$. This representation depends on the choice of bases of V and W, and it is possible that expansions with respect to two different sets of bases might different in length.

Definition 7.7. Let V and W be vector spaces over a field F and let $(V \otimes W, \varphi)$ be the tensor product of V and W. The *rank* of a nonzero vector $x \in V \otimes W$ is the largest positive integer N such that, if

$$x = \sum_{k=1}^{n} v_k \otimes w_k$$

for some $v_1, \ldots, v_n \in V \setminus \{0\}$ and $w_1, \ldots, w_n \in W \setminus \{0\}$, then $n \geq N$. If $N = 1$, then x is said to be an *elementary tensor.*

We have not defined *zero-rank tensors*, as the only reasonable candidate would be the zero vector. When is $\sum_{k=1}^{n} v_k \otimes w_k = 0$? The following theorem provides the answer.

Theorem 7.8. *Let V and W be vector spaces over a field F and let $(V \otimes W, \varphi)$ be the tensor product of V and W. Let $v_1, \ldots, v_n \in V$ and $w_1, \ldots, w_n \in W$. The following are equivalent:*

(1) $\sum_{k=1}^{n} v_k \otimes w_k = 0$;
(2) $\sum_{k=1}^{n} l(v_k)u(w_k) = 0$ for all $l \in V^\star$ and $u \in W^\star$;
(3) $\sum_{k=1}^{n} l(v_k)w_k = 0$ for all $l \in V^\star$;
(4) $\sum_{k=1}^{n} u(w_k)l(v_k) = 0$ for all $u \in W^\star$.

Proof of theorem. (1) ⇒ (2) For each $l \in V^\star$ and every $u \in W^\star$, we define a bilinear map $(l, u) : V \times W \to F$ by setting $(l, u)(v, w) = l(v)u(w)$ for all $v \in V$ and $w \in V$. Let $\widetilde{(l, u)} : V \otimes W \to F$ denote the mediating morphism of (l, u) and observe that

$$\begin{aligned} 0 = \widetilde{(l, u)} \left(\sum_{k=1}^{n} v_k \otimes w_k \right) &= \sum_{k=1}^{n} \widetilde{(l, u)}(v_k \otimes w_k) = \sum_{k=1}^{n} \left(\widetilde{(l, u)} \circ \varphi \right) (v_k, w_k) \\ &= \sum_{k=1}^{n} (l, u)(v_k, w_k) = \sum_{k=1}^{n} l(v_k)u(w_k), \end{aligned}$$

as was to be shown.

(2) ⇒ (1). For each $l \in V^\star$ and every $u \in W^\star$, we define a bilinear map $(l, u) : V \times W \to F$ by setting $(l, u)(v, w) = l(v)u(w)$ for all $v \in V$ and $w \in W$. Let

$\widetilde{(l,u)} : V \otimes W \to F$ denote the mediating morphism of (l,u) and observe that

$$\begin{aligned} 0 &= \sum_{k=1}^{n} l(v_k)u(w_k) = \sum_{k=1}^{n} (l,u)(v_k,w_k) = \sum_{k=1}^{n} \left(\widetilde{(l,u)} \circ \varphi\right)(v_k,w_k) \\ &= \sum_{k=1}^{n} \widetilde{(l,u)}(v_k \otimes w_k) = \widetilde{(l,u)}\left(\sum_{k=1}^{n} v_k \otimes w_k\right). \end{aligned} \tag{7.9}$$

Let $x = \sum_{k=1}^{n} v_k \otimes w_k$ and suppose for a contradiction that $x \neq 0$. Let N be the rank of x and find $v_1', \ldots, v_N' \in V$ and $w_1', \ldots, w_N' \in W$ such that

$$x = \sum_{k=1}^{N} v_k' \otimes w_k'.$$

It follows from bilinearity that $\{v_1', \ldots, v_N'\}$ and $\{w_1', \ldots, w_N'\}$ are linearly independent subsets of V and W, respectively. Computations analogous to (7.9) yield the following string of identities for all $l \in V^\star$ and $u \in W^\star$.

$$0 = \widetilde{(l,u)}\left(\sum_{k=1}^{N} v_k' \otimes w_k'\right) = \sum_{k=1}^{N} l(v_k')u(w_k').$$

Since $v_1', \ldots, v_N'$ are linearly independent, we can take the dual vectors $(v_1')^\star, \ldots,$ $(v_N')^\star$ (Corollary 6.6) and substitute them in place of l in the above display to obtain the following:

$$0 = \sum_{k=1}^{N} (v_i')^\star u(w_k') = u(w_i') \text{ for all } 1 \leq i \leq N \text{ and } u \in W^\star.$$

It follows that $w_i' = 0$ for all $1 \leq i \leq N$, for otherwise we can invoke the algebraic Hahn–Banach theorem (Lemma 6.9) to construct a $u \in W^\star$ such that $u(w_i') \neq 0$. We now conclude from bilinearity that $x = \sum_{k=1}^{N} v_k' \otimes w_k'$ is the zero vector.

(2) $\Rightarrow$ (3). Since $u\left(\sum_{k=1}^{n} l(v_k)w_k\right) = 0$ for all $l \in V^\star$, it follows from the algebraic Hahn–Banach theorem (Lemma 6.9) that $\sum_{k=1}^{n} l(v_k)w_k = 0$.

(3) $\Rightarrow$ (2). Since $\sum_{k=1}^{n} l(v_k)w_k = 0$, we see that

$$\sum_{k=1}^{n} l(v_k)u(w_k) = u\left(\sum_{k=1}^{n} l(v_k)w_k\right) = 0.$$

(2) $\Rightarrow$ (4). Analogous to (2) $\Rightarrow$ (3).

(4) $\Rightarrow$ (2). Analogous to (3) $\Rightarrow$ (2). □

The proof of (2) $\Rightarrow$ (1) suggests the following corollary:

Corollary 7.10. *If $v_1, \ldots, v_n$ are linearly independent vectors in V and $w_1, \ldots, w_n$ are vectors in W, then $\sum_{k=1}^{n} v_k \otimes w_k = 0$ implies that $w_k = 0$ for all $1 \leq k \leq n$. In particular, $v \otimes w = 0$ if and only if $v = 0$ and $w = 0$.*

The last part of the proof of (2) $\Rightarrow$ (1) can be used almost verbatim to establish the corollary. Here we give an alternate proof.

Proof. Theorem 7.8 implies that

$$0 = \sum_{k=1}^{n} u(w_k)v_k$$

for all $u \in W^\star$. Since $v_1, \ldots, v_n$ are linearly independent, we see that $u(w_k) = 0$ for all $1 \leq k \leq n$. Since $u \in W^\star$ was arbitrary, it follows from the algebraic Hahn–Banach theorem (Lemma 6.9) that $w_k = 0$ for all $1 \leq k \leq n$. $\square$

Another interesting point about the proof of (2) $\Rightarrow$ (1) is that, if $\widetilde{(l, u)}$ in (7.9) accounted for all linear functionals on $V \otimes W$, then we would have been able to invoke the algebraic Hahn–Banach theorem (Lemma 6.9) to conclude that $\sum_{k=1}^n v_k \otimes w_k = 0$. To what extent is this argument valid? The following theorem provides an answer.

Theorem 7.11 (Tensor product of linear functionals)**.** *Let V and W be vector spaces over a field F. There exists a unique injective linear operator $\theta : V^\star \otimes W^\star \to (V \otimes W)^\star$ defined by the formula $\theta(l \otimes u) = l \otimes u$, where*

$$(l \otimes u)(v \otimes w) = l(v)u(w).$$

If $\dim V < \infty$ and $\dim W < \infty$, then θ is an isomorphism.

Proof. The map $\Theta : V^\star \times W^\star \to (V \otimes W)^\star$ defined by the formula $\Theta(l, u) = l \otimes u$ is bilinear. The mediating morphism $\theta : V^\star \otimes W^\star \to (V \otimes W)^\star$ is the desired map, whose uniqueness is guaranteed by the uniqueness clause in the construction of the tensor product (Theorem 7.5).

To show that θ is injective, we fix $x \in \ker \theta$. Suppose for a contradiction that $x \neq 0$. Let N be the rank of x and take an expansion $x = \sum_{k=1}^N l_k \otimes u_k$. By bilinearity, $\{l_1, \ldots, l_N\}$ and $\{u_1, \ldots, u_N\}$ are linearly independent subsets of $V^\star$ and $W^\star$, respectively.

Observe that

$$0 = \theta\left(\sum_{k=1}^N l_k \otimes u_k\right)(v \otimes w) = \sum_{k=1}^N l_k(v)u_k(w).$$

for all $v \in V$ and $w \in W$. We fix $v \in V \smallsetminus \{0\}$ and note that $w \mapsto \sum_{k=1}^N l_k(v)u_k(w)$ is the zero functional. Since $w_1, \ldots, w_N$ are linearly independent, we conclude that $l_1(v) = \cdots = l_N(v) = 0$. The choice of v was arbitrary, and so $l_1 = \cdots = l_N$. It follows that $x = 0$, which is absurd. We now see that $\ker \theta$ is trivial.

Finally, if V and W are finite-dimensional, then Corollary 6.6 and Theorem 7.6 imply that

$$\begin{aligned}\dim(V^\star \otimes W^\star) &= (\dim V^\star)(\dim W^\star) = (\dim V)(\dim W) \\ &= \dim(V \otimes W) = \dim\left((V \otimes W)^\star\right).\end{aligned}$$

Since the first isomorphism theorem implies that $(V^\star \otimes W^\star) \cong \operatorname{im} \theta$, we conclude from the above dimension computations that $(V^\star \otimes W^\star) \cong \operatorname{im} \theta = (V \otimes W)^\star$. $\square$

We now generalize the above theorem to characterize tensor products of linear transformations

Theorem 7.12 (Tensor product of linear transformations)**.** *Let V_1, V_2, W_1, W_2 be vector spaces over a field F. There exists a unique injective linear operator $\theta : \mathscr{L}(V_1, V_2) \otimes \mathscr{L}(W_1, W_2) \to \mathscr{L}(V_1 \otimes W_1, V_2 \otimes W_2)$ defined by the formula $\theta(T, S) = T \otimes S$, where*

$$(T \otimes S)(v \otimes w) = Tv \otimes Sw.$$

If V_1, V_2, W_1, W_2 are finite-dimensional, then θ is an isomorphism.

Proof. The map $\Theta : \mathscr{L}(V_1, V_2) \times \mathscr{L}(W_1, W_2) \to \mathscr{L}(V_1 \otimes W_1, V_2 \otimes W_2)$ defined by the formula $\Theta(T, S) = T \otimes S$ is bilinear. The mediating morphism $\theta : \mathscr{L}(V_1, V_2) \otimes \mathscr{L}(W_1, W_2) \to \mathscr{L}(V_1 \otimes W_1, V_2 \otimes W_2)$ is the desired map, whose uniqueness is guaranteed by the uniqueness clause in the construction of the tensor product (Theorem 7.5).

To show that θ is injective, we fix $U \in \ker \theta$. Suppose for a contradiction that $U \neq 0$. Let N be the rank of U and take an expansion $U = \sum_{k=1}^{N} T_k \otimes S_k$. By bilinearity, $\{T_1, \ldots, T_N\}$ and $\{S_1, \ldots, S_N\}$ are linearly independent subsets of $\mathscr{L}(V_1, V_2)$ and $\mathscr{L}(W_1, W_2)$, respectively.

Observe that

$$0 = \theta \left(\sum_{k=1}^{N} T_k \otimes S_k \right) (v \otimes w) = \sum_{k=1}^{N} T_k v \otimes S_k w$$

for all $v \in V$ and $w \in W$. We fix $v \in V$ and observe that $w \mapsto \sum_{k=1}^{N} T_k v \otimes S_k w$ is the zero map. Since $w_1, \ldots, w_N$ are linearly independent, Corollary 7.10 implies that $T_1 v = \cdots = T_N v = 0$. The choice of v was arbitrary, and so $T_1, \ldots, T_N$ are zero maps. It follows that $U = 0$, which is absurd. We conclude that $\ker \theta$ is trivial.

The isomorphism statement follows from the usual dimension-counting argument: see, for example, the proof of Theorem 7.11. $\square$

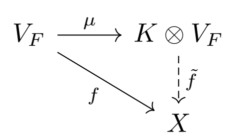
We conclude this section with a discussion on how to change the base field of a vector space. If K is a field extension of a field F, then a vector space V_K over K can be thought of as a vector space V_F. We tackle the converse question: given a vector space V_F, how can we turn V_F into a vector space over K?

Consider the tensor product $W_F = K \otimes V_F$. Since K can be thought of as an F-vector space, W_F is a well-defined tensor product of two F-vector spaces. On the other hand, we can think of $K \otimes V_F$ as a K-vector space by setting

$$a(b \otimes v) := (ab) \otimes v$$

for all $a, b \in K$ and $v \in V_F$. We then obtain the K-vector space $W_K = K \otimes V_F$, which has the same elements as W_F but is a vector space over a different field.

Theorem 7.13. *Let K be a field extension of a field F and let V_F be a vector space over F. Define the* K-extension map *$\mu : V_F \to K \otimes V_F$ by setting $\mu(v) = 1_K \otimes v$ for all $v \in V_F$. The map μ is F-linear, and, for each F-linear map $f : V_F \to X$ where X is a K-vector space, there exists a unique K-linear map $\tilde{f} : K \otimes V_F \to X$ such that the following diagram commutes:*



Proof. Linearity of μ follows at once from bilinearity of $\otimes$. Fix an F-linear map $f : V_F \to X$ where X is a K-vector space. The map $\tilde{f} : K \otimes V_F \to X$ given by the formula $\tilde{f}(a \otimes v) = af(v)$ is the unique K-linear map such that the diagram commutes. $\square$

Corollary 7.14. *Let K be a field extension of a field F, let V and W be vector spaces over F, and let $\mu_V : V \to K \otimes V$ and $\mu_W : W \to K \otimes W$ be the K-extension* *maps. For each F-linear map $f : V \to W$, there exists a unique K-linear map $\tilde{f} : K \otimes V \to K \otimes W$ such that the following diagram commutes:*

$$\begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ {\scriptstyle \mu_V}\big\downarrow & & \big\downarrow{\scriptstyle \mu_W} \\ K \otimes W & \xrightarrow[\tilde{f}]{} & K \otimes W \end{array}$$

Proof. Existence and uniqueness follows at once from Theorem 7.13. In fact, $\mathrm{id}_K \otimes f : K \otimes V \to K \otimes W$ is the unique such map. □

8. Normed Linear Spaces

We now turn to the analytic properties of vector spaces. A standard object of study in the analysis on vector spaces is a *normed linear space*:

Definition 8.1. A *norm* on a vector space X over $\mathbb{F}$[1] is the function $\|\cdot\| : X \to [0, \infty)$ satisfying the following properties:

(i) $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in X$;
(iii) $\|\lambda x\| = |\lambda|\|x\|$ for each $\lambda \in \mathbb{F}$ and every $x \in X$.

X is a *normed linear space* if X is equipped with a norm.

With every normed linear space is a metric topology, given by the induced metric

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

Remark 8.2. This topology turns vector addition $(x, y) \mapsto x + y$ and scalar multiplication $(\lambda, x) \mapsto \lambda x$ into continous maps.

The induced metric satisfies the following properties:

(i) **Translation invariance**. $d(x + z, y + z) = d(x, y)$ for all $x, y, z \in X$.
(ii) **Homogeneous scaling**. $d(\lambda x, \lambda y) = |\lambda| d(x, y)$ for each $\lambda \in \mathbb{F}$ and every $x, y \in X$.

Remark 8.3. Every vector space X with a metric that satisfies (i) and (ii) is a normed vector space with the norm given by $\|x\| = d(x, 0)$.

Translation invariance and homogeneous scaling of the induced metric has the following consequences:

Proposition 8.4. *Let X be a normed linear space over $\mathbb{F}$.*

(i) $\overline{A + y} = \overline{A} + y$ and $(A + y)^\circ = A^\circ + y$ for each $A \subseteq X$ and every $y \in X$.
(ii) $\overline{\lambda A} = \lambda \overline{A}$ and $(\lambda A)^\circ = \lambda A^\circ$ for each $\lambda \in \mathbb{F}$ and every $A \subseteq X$.
(iii) If U is open subset of X and A an arbitrary subset of X, then $U + A$ is open.
(iv) If A and B are compact subsets of X, then $A + B$ is compact.

We remark that the sum of two closed subsets is not necessarily closed. Indeed, $\mathbb{Z} + \sqrt{2}\mathbb{Z}$ is a proper dense subset of $\mathbb{R}$, which cannot be closed.

Another basic property of the norm topology is that *the norm is continuous.* In fact, the map $x \mapsto \|x\|$ is Lipschitz, for the triangle inequality yields the estimate

$$|\|x\| - \|y\|| \leq \|x - y\|.$$

[1] Henceforth, $\mathbb{F}$ shall denote an "arbitrary field": $\mathbb{R}$ or $\mathbb{C}$.

As a consequence, the *open ball*

$$B_r(x) = \{y : \|x - y\| < r\}$$

is open, and the *closed ball*

$$B_r[x] = \{y : \|x - y\| \leq r\}$$

is closed. Translation and scaling yields

$$B_r(x) = x + rB_1(0) \quad \text{and} \quad B_r[x] = x + rB_1[0].$$

Remark 8.5. A *metric linear space* is a vector space with a translation-invariant metric. Every ball in a normed linear space is convex, but $\mathbb{R}^2$ with the metric

$$d(x, y) = \sqrt{|x_1 - y_1|} + \sqrt{|x_2 - y_2|}$$

is a metric linear space whose unit ball $B_1(0)$ fails to be convex.

A basic result in the theory of metric spaces is that $\overline{B_r(x)} \subseteq B_r[x]$ and $B_r[x]^\circ \supseteq B_r(x)$. In a general metric space, the equalities need not hold, for any set of cardinality at least 2 with the discrete metric

$$d(x, y) = \begin{cases} 0 & \text{if } x = y; \\ 1 & \text{if } x \neq y; \end{cases}$$

has the following "unit balls":

$$B_1(x) = \{x\}, \quad \overline{B_1(x)} = \{x\}, \quad B_1[x] = X.$$

The equality holds in normed linear spaces:

Proposition 8.6. *In a normed linear space,* $\overline{B_1(0)} = B_1[0]$.

Proof. It suffices to show that $B_1[0] \subseteq \overline{B_1(0)}$. We pick $x \in B_1[0]$ and construct a sequence

$$x_n = \left(1 - \frac{1}{n}\right) x.$$

Then $x_n \in B_1(0)$ for each $n \in \mathbb{N}$ and $x_n \to x$ in the norm topology, whence $x \in \overline{B_1(0)}$. □

Let us now introduce some fundamental examples of normed linear spaces.

Example 8.7. The *p-norms* on $\mathbb{F}^n$, given by

$$\|x\|_p = \begin{cases} \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p} & \text{if } 1 \leq p < \infty; \\ \max_i |x_i| & \text{if } p = \infty; \end{cases}$$

is a norm. More generally, if (X, μ) is a measure space, then the *Lebesgue space* $L^p(X, \mu)$ of μ-measurable complex-valued functions on X with finite L^p-norm

$$\|f\|_p = \begin{cases} \left(\int_X |f|^p \, d\mu\right)^{1/p} & \text{if } 1 \leq p < \infty; \\ \operatorname{ess\,sup} |f| & \text{if } p = \infty; \end{cases}$$

is a normed linear space.

Example 8.8. Let E be an arbitrary set, and consider the collection $\mathscr{B}(E)$ of bounded real-valued functions on E. The L^∞-norm is a norm on $\mathscr{B}(E)$. In particular, if E is a compact metric space, then the space $\mathscr{C}(E)$ of continuous real-valued functions on E with the L^∞-norm is a normed linear space. Similarly, $C^1([0,1])$ of continuously differentiable functions on $[0,1]$ with a "modified" L^∞-norm

$$\|f\| = \|f\|_\infty + \|f'\|_\infty$$

is a normed linear space.

Example 8.9. If A is an invertible n-by-n matrix over $\mathbb{F}$, then

$$\|x\|_{A,p} = \|Ax\|_p$$

is a norm on $\mathbb{F}^n$.

The Lebesgue spaces are more than just normed linear spaces— they are complete metric spaces. Since these examples are of paramount importance in analysis, we abstract this property and give it the name of one of the founders of functional analysis, Stefan Banach:

Definition 8.10. A *Banach space* is a normed linear space whose induced metric is complete.

Not every space is a Banach space, of course. $\mathscr{C}([0,1])$, as a subset of $L^2([0,1])$, is a dense proper subspace, hence it cannot be closed. Nevertheless, there is a Banach space that contains it: namely, $L^2([0,1])$. As it turns out, we can always find a Banach space that contains the given normed linear space.

For a precise formulation of this result, we will need some definitions. Recall that an *isomorphism* of two vector spaces is a bijective linear transformation. An isomorphism of vector spaces preserves linear structures in a way that renders the two spaces in question "essentially the same". An appropriate analogue for normed linear spaces is as follows:

Definition 8.11. An *isometric isomorphism* between normed linear spaces X and Y is an isomorphism $T : X \to Y$ of vector spaces that is *isometric*, viz., $\|x\| = \|Tx\|$ for all $x \in X$. X and Y are said to be *isometrically isomorphic* if there exists an isometric isomorphism between them.

The adjective *isometric* is important, because there is another notion of isomorphism for normed linear spaces: namely, bijective linear transformations that are homeomorphisms with respect to the metric topology generated by the norm. See §17.3 for a more detailed discussion.

Isometrically isomorphic normed linear spaces are considered to be the "same", and so the result we wish to formalize must be understood in the isomorphic sense as well:

Definition 8.12. An *embedding* of a vector space X into another vector space Y is an injective linear transformation $T : X \to Y$. An embedding $T : X \to Y$ is said to be an *isometric embedding* if X and Y are normed linear spaces and T is an isometry.

It is easy to see that T is an isometric isomorphism between X and $\operatorname{im} T$. We are now in a position to state the following theorem:

Theorem 8.13. *Every normed linear space can be isometrically embedded into a Banach space. The Banach space can be chosen in a way that the embedding is dense. This choice is unique up to an isometric isomorphism and is referred to as the* completion *of the normed linear space in question.*

We will prove this result in §11 of this chapter. For now, we consider the following simple exercise that highlights the basic concepts related to the above result.

Remark 8.14. Let $(X, \|\cdot\|)$ be a normed linear space and M a linear subspace of X. This this case, $\overline{M}$ is also a linear subspace of X. Moreover, if $(x_n)_{n=1}^\infty$ is a sequence in X, then $x_n \to x$ implies $\|x_n\| \to \|x\|$ in $\mathbb{R}$. If $(x_n)_{n=1}^\infty$ is a Cauchy sequence, then $(\|x_n\|)_{n=1}^\infty$ is a Cauchy sequence.

Before we proceed, we remark that not every metric on a vector space is induced by a norm.

Remark 8.15. The metric

$$d(x,y) = \sum_{n=1}^\infty 2^{-n} \min(1, |x_n - y_n|)$$

is a translation-invariant metric on $\mathbb{R}^{\mathbb{N}}$ that does not scale homogeneously. This is an example of a *Fréchet space* (Definition 15.7), a generalization of a Banach space with multiple, rather than one, "complete norms".

We also remark that not every norm generates a distinct norm topology.

Definition 8.16. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on X are *equivalent* if there exist two positive constants K_1 and K_2 such that

$$K_1\|x\|_a \leq \|x\|_b \leq K_2\|x\|_a$$

for all $x \in X$.

Remark 8.17. Equivalent norms induce the same metric topology, viz., a subset E is open with respect to $\|\cdot\|_a$ if and only if E is open with respect to $\|\cdot\|_b$.

It turns out that there is only one norm topology on finite-dimensional vector spaces of a fixed dimension.

Theorem 8.18. *All norms on a finite-dimensional vector space are equivalent.*

Proof. Let X be a finite-dimensional normed linear space over $\mathbb{F}$ with the norm $\|\cdot\|$. We fix a basis $\{v_1, \ldots, v_n\}$ of X and consider the canonical isomorphism $T : \mathbb{F}^n \to X$, given by

$$T(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^n \lambda_i v_i$$

for each $\lambda = (\lambda_1, \ldots, \lambda_n)$ in $\mathbb{F}^n$. The Cauchy-Schwarz inequality implies that

$$\|T\lambda\| \leq \sum_{i=1}^n |\lambda_i|\|v_i\| \leq \left(\sum_{i=1}^n \|v_i\|^2\right)^{1/2} \left(\sum_{i=1}^n |\lambda_i|^2\right)^{1/2} = K_2\|\lambda\|_2$$

for each $\lambda \in \mathbb{F}^n$. From this, we see that $\lambda \mapsto \|T\lambda\|$ is continuous, for

$$|\|T\lambda_1\| - \|T\lambda_2\|| \leq \|T(\lambda_1 - \lambda_2)\| \leq K_2\|\lambda_1 - \lambda_2\|_2.$$

We now take the unit sphere $S = \{x \in \mathbb{F}^n : \|x\|_2 = 1\}$ in $\mathbb{F}^n$, which is compact. The map $\lambda \mapsto \|T\lambda\|$ therefore attains its minimum K_1 on S. The minimum must be strictly postiive, for the norm is only zero at 0. We then have

$$\|T\lambda\| = \|\lambda\|\|T(\|\lambda\|^{-1}\lambda)\| \geq K_1\|\lambda\|_2$$

for all nonzero $\lambda \in \mathbb{F}^n$, and it follows that

$$K_1\|\lambda\|_2 \leq \|T\lambda\| \leq K_2\|\lambda\|_2. \tag{8.19}$$

We now take another norm $\|\cdot\|'$ on X. The same argument as above shows that we have positive constants K_1' and K_2' such that

$$K_1'\|\lambda\|_2 \leq \|T\lambda\|' \leq K_2'\|\lambda\|_2. \tag{8.20}$$

Let $x = T\lambda$. We apply (8.20) and then (8.19) to obtain the estimate

$$\|x\|' \leq K_2'\|\lambda\|_2 \leq \frac{K_2'}{K_1}\|x\|.$$

Similarly, we apply (8.19) and then (8.20) to obtain the estimate

$$\|x\| \leq K_2\|\lambda\|_2 \leq \frac{K_2}{K_1'}\|x\|'.$$

Combining the two estimates, we have

$$\frac{K_1'}{K_2}\|x\|' \leq \|x\| \leq \frac{K_2'}{K_1}\|x\|'$$

for all $x \in \operatorname{im} T$. Since T is an isomorphism, this accounts for all vectors in X. □

The equivalence of norms implies that all finite-dimensional normed linear spaces of the same dimension are homeomorphic. Since finite-dimensional vector spaces of the same dimension are isomorphic, we see that they are isomorphic as topological vector spaces. They are not necessarily isomorphic as normed linear spaces, however.

Example 8.21. Consider $(\mathbb{R}^2, \|\cdot\|_1)$ and $(\mathbb{R}^2, \|\cdot\|_2)$. If $T : (\mathbb{R}^2, \|\cdot\|_1) \to (\mathbb{R}^2, \|\cdot\|_2)$ is an isometric isomorphism, then

$$\|(1-\lambda)x + \lambda y\|_1 = \|(1-\lambda)Tx + \lambda Ty\|_2 \tag{8.22}$$

for all $\lambda \in \mathbb{R}$. Letting $x = (1, 0)$ and $y = (0, 1)$, we see that

$$\|(1-\lambda)x + \lambda y\|_1 = 1$$

for each $0 \leq \lambda \leq 1$. But Tx and Ty are points on the ordinary unit circle in the coordinate plane $\mathbb{R}^2$, whence the segment connecting the two points cannot be on the unit circle. It follows that

$$\|(1-\lambda)x + \lambda y\|_2 < 1$$

if $0 < \lambda < 1$, contradicting (8.22). It follows that there is no isometric isomorphism from $(\mathbb{R}^2, \|\cdot\|_1)$ to $(\mathbb{R}^2, \|\cdot\|_2)$.

Nevertheless, all finite-dimensional normed linear spaces of the same dimension are similar enough for the most part. For one, all of them are complete:

Corollary 8.23. *All finite-dimensional normed linear spaces are Banach spaces.*

Proof. Let $(X, \|\cdot\|)$ be an n-dimensional normed linear space and fix a Cauchy sequence $(x_n)_{n=1}^\infty$ in V. The equivalence of norms (Theorem 8.18) implies that the Euclidean norm (l^2-norm) on $\mathbb{R}^n$ and $\|\cdot\|$ are equivalent. More precisely, if we fix a basis $\{v_1, \ldots, v_n\}$ of X and take the canonical isomorphism $T : X \to \mathbb{R}^n$, given by

$$T\left(\sum_{i=1}^n \lambda_i v_i\right) = (\lambda_1, \ldots, \lambda_n),$$

then we can find positive constants K_1 and K_2 such that

$$K_1 \|Tx\|_2 \leq \|x\| \leq K_2 \|Tx\|_2. \tag{8.24}$$

We now note that $(Tx_n)_{n=1}^\infty$ is Cauchy in $\mathbb{R}^n$, and so we can find its limit y. We set $x = T^{-1}y$, so that $\|Tx_n - Tx\|_2 \to 0$ as $n \to \infty$. (8.18) now implies that

$$\lim_{n\to\infty} \|x_n - x\| \leq \lim_{n\to\infty} K_2 \|Tx_n - Tx\|_2 = 0,$$

whence X is a Banach space. □

This, in particular, shows that all finite-dimensional subspaces of a normed linear space are complete. Since completeness implies closedness, we have the following corollary:

Corollary 8.25. *All finite-dimensional subspaces of a normed linear space are closed.* □

We remark that the finiteness assumption is essential. Indeed, there are many examples of infinite-dimensional subspaces that are not closed.

Remark 8.26. Fix $1 \leq p < \infty$ and recall that the space $\mathcal{C}^\infty([0,1])$ of smooth functions is a dense subspace of $L^p([0,1])$. Therefore, the space $\mathcal{C}^n([0,1])$ of n-times continuously differentiable functions on $[0,1]$ is not closed in the norm topology of $L^p([0,1])$.

Remark 8.27. Nevertheless, there are concrete examples of nontrivial closed infinite-dimensional subspaces of infinite-dimensional spaces. For example, the space of bounded and continuous functions on $\mathbb{R}$ is a proper closed subspace of $L^\infty(\mathbb{R})$.

We also note that finiteness is used in a crucial manner in the proof of the equivalence of norms: first, to compute the constant $(\sum_{i=1}^n \|v_i\|^2)^{1/2}$, and second, to obtain the minimum of the map $\lambda \mapsto \|T\lambda\|$ via the compactness of the unit sphere in $\mathbb{F}^n$.

Example 8.28. On an infinite-dimensional vector space, we can easily construct non-equivalent norms. Consider, for example, the vector space X of sequences $(x_n)_{n=1}^\infty$ that vanish after finitely many terms. X can then be considered as linear subspaces of $l^1(\mathbb{N})$ and $l^2(\mathbb{N})$, respectively. We now let

$$x_n^N = \begin{cases} 1 & \text{if } n \leq N; \\ 0 & \text{if } n > N; \end{cases}$$

and consider the family of sequences $x^N = (x_n^N)_{n=1}^\infty$. We have

$$\|x^N\|_1 = N \quad \text{and} \quad \|x^N\|_2 = \sqrt{N},$$

and, by sending $N \to \infty$, we see that no norm equivalence is possible.

More generally, every infinite-dimensional vector space admits non-equivalent norms.

Remark 8.29. Let V be an infinite-dimensional vector space and let $\mathcal{B} = \{v_\alpha\}$ be a Hamel basis of V. Check that each function $f : \mathcal{B} \to [0, \infty)$ defines a norm $\|\cdot\|_f$, given by

$$\left\| \sum_\alpha \lambda_\alpha v_\alpha \right\| = \sum_\alpha |\lambda_\alpha| f(v_\alpha).$$

Conclude that V admits non-equivalent norms $\|\cdot\|_f$ and $\|\cdot\|_g$.

This suggests that the unit sphere in an infinite-dimensional normed linear space is not compact, for otherwise the proof of Theorem 8.18 would have gone through.

Theorem 8.30. *A normed linear space is finite-dimensional if and only if the closure of each bounded subset is compact.*

Remark 8.31 ([Tao10], Example 1.9.8). As a concrete example, we consider $l^p(\mathbb{N})$ with $1 \leq p < \infty$. The set $\{e_n\}_{n\in\mathbb{N}}$ of sequences

$$e_n(m) = \begin{cases} 1 & \text{if } n = m; \\ 0 & \text{if } n \neq m; \end{cases}$$

is closed and bounded in $l^p(\mathbb{N})$, but is not compact. Nevertheless, a closed and bounded set $K \subseteq l^p(\mathbb{N})$ such that every $\varepsilon > 0$ furnishes an index n with the estimate

$$\left(\sum_{m>n} |f(m)|^p \right)^{1/p} \leq \varepsilon$$

for all $f \in K$ is compact.

Proof of theorem. The "only if" part follows at once from the equivalence of norms. To show the "if" part, we assume that X is infinite-dimensional. We shall need the following lemma, due to F. Riesz:

Lemma 8.32 (Riesz)**.** *Let $(X, \|\cdot\|)$ be a normed linear space and Y a closed, proper subspace of X. For each $\varepsilon > 0$, we can find a vector $z \in X$ such that $\|z\| = 1$ and $\|z - y\| > 1 - \varepsilon$ for all $y \in Y$.*

Proof of lemma. $X \smallsetminus Y$ is nontrivial, hence we can pick a vector x_0 in it. The quantity

$$d = \inf_{y \in Y} \|x_0 - y\|$$

is positive, for otherwise x_0 is in the closure of Y. Since Y is closed, this is obviously false.

Now, for each $\eta > 0$, we can find $y_\eta \in Y$ such that

$$d \leq \|x_0 - y_\eta\| < d + \eta.$$

We let

$$z_\eta = \|x_0 - y_\eta\|^{-1}(x_0 - y_\eta).$$

For each $y \in Y$, we have the estimate

$$\begin{aligned} \|z_\eta - y\| &= \left\|y - \|x_0 - y_\eta\|^{-1}(x_0 - y_\eta)\right\| \\ &= \frac{1}{\|x_0 - y_\eta\|} \left\| \|x_0 - y_\eta\| y - (x_0 - y_\eta)\right\| \\ &= \frac{1}{\|x_0 - y_\eta\|} \|x_0 - (y_\eta + \|x_0 - y_\eta\| y)\| \\ &> \frac{1}{d+\eta} \cdot d, \end{aligned}$$

because $y_\eta + \|x_0 - y_\eta\| y \in Y$.

We now fix a sufficiently small $\varepsilon > 0$ and pick η such that $\frac{d}{d+\eta} = 1 - \varepsilon$. Letting $z = z_\eta$, we see that the above estimate translates to $\|z\| = 1$ and

$$\|z - y\| > \frac{d}{d+\eta} = 1 - \varepsilon,$$

which was the conclusion of the lemma. □

We now suppose for a contradiction that the closure of each bounded subset of X is compact. Fix a unit vector x_1 in X and let Y_1 be the span of x_1. The lemma above furnishes a unit vector x_2 in X such that $\|x_2 - y\| \geq \frac{1}{2}$ for all $y \in Y_1$. In particular, $\|x_2 - x_1\| \geq \frac{1}{2}$, and so the span Y_2 of x_1 and x_2 is distinct from Y_1. Since X is infinite-dimensional, we continue this process to obtain a strictly increasing sequence

$$Y_1 \subset Y_2 \subset \cdots \subset Y_n \subset \cdots .$$

of subspaces of X, and the corresponding sequence $(x_n)_{n=1}^\infty$ of vectors such that $\|x_m - x_n\| \geq \frac{1}{2}$ whenever $m \neq n$. We can see at once that the sequence $(x_n)_{n=1}^\infty$ has no convergent subsequence, despite its boundedness. It follows that $\{x_n\}_{n \in \mathbb{N}}$ is a closed, bounded subset of X that is not compact, which is absurd. □

9. Bounded Linear Operators

The argument for the continuity of $\lambda \mapsto \|T\lambda\|$ given in proof of the equivalence of norms in finite-dimensional vector spaces (Theorem 8.18) shows that $\lambda \mapsto T\lambda$ is Lipschitz continuous. Indeed, we have

$$\|T\lambda_1 - T\lambda_2\| \leq C\|\lambda_1 - \lambda_2\|_2$$

with the positive constant $C = (\sum_{i=1}^n \|v_i\|^2)^{1/2}$, and the Lipschitz continuity follows at once. Again, finiteness plays an essential role here, as not every linear transformation between infinite-dimensional normed linear spaces is continuous.

Example 9.1. For example, we consider the differential operator $D : \mathscr{C}^1([0,1]) \to \mathscr{C}([0,1])$ given by

$$Df = f'.$$

We let $f_n(x) = n^{-1}\sin(nx)$ for each $n \in \mathbb{N}$ and observe that $\|f_n\|_\infty \to 0$ in $\mathcal{C}^1[0,1]$. We see, however, that $Df_n = \cos nx$, and so $\|Df_n\|_\infty = 1$ for all $n \in \mathbb{N}$. Therefore, $Df_n \not\to 0$ in $\mathcal{C}[0,1]$, whence D cannot be continuous.

Example 9.2. As another example, we consider the space X of sequences that vanish after finitely many terms. The map

$$l(x) = \sum_{n=1}^{\infty} x_n$$

is a linear transformation from X to $\mathbb{R}$. We now let

$$x_n^N = \begin{cases} \frac{1}{N} & \text{if } n \leq N; \\ 0 & \text{if } n > N; \end{cases}$$

for each $N \in \mathbb{N}$ and observe that $\|x^N\|_\infty \to 0$ in X. Nevertheless, $l(x^N) = 1$ for all $N \in \mathbb{N}$, so that $l(x^N) \not\to 0$ in $\mathbb{R}$. It follows that l cannot be continuous.

If, however, we can find a positive constant C such that

$$\|Tx\| \leq C\|x\|,$$

as we have in the proof of the equivalence of norms, then it is straightforward to show that T is, in fact, continuous. We give this condition a name.

Definition 9.3. A linear transformation $T : X \to Y$ between $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ is *bounded* if there exists a positive constant C such that

$$\|Tx\| \leq C\|x\|$$

for all $x \in X$.

As it turns out, boundendess is also necessary for a linear transformation to be continuous, rendering boundedness an equivalent condition for continuity.

Proposition 9.4. *Let $T : X \to Y$ be a linear transformation between $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$. The following are equivalent:*

(a) T is uniformly continuous.
(b) T is continuous.
(c) T is continuous at 0.
(d) T is bounded.

Proof. (a) $\Rightarrow$ (b) and (b) $\Rightarrow$ (c) are trivial. To see (c) $\Rightarrow$ (a), we fix $\varepsilon > 0$ and find a $\delta > 0$ such that $\|x\|_X < \delta$ implies $\|Tx\|_Y < \varepsilon$. Let $x_1, x_2 \in X$. If $\|x_1 - x_2\|_X < \delta$, then $\|T(x_1 - x_2)\|_Y = \|Tx_1 - Tx_2\|_Y < \varepsilon$, and so T is uniformly continuous.

(d) $\Rightarrow$ (c) is straightforward, hence it remains to show that (c) $\Rightarrow$ (d). To this end, we find a $\delta > 0$ such that $\|x\| < \delta$ implies $\|Tx\| = \|Tx - T(0)\| < 1$. Given an arbitrary vector $x \in X$, we now see that

$$\|Tx\| = \frac{2\|x\|}{\delta} \left\| T\left(\frac{\delta}{2\|x\|} x\right) \right\| < \frac{2}{\delta}\|x\|,$$

and so T is bounded. □

We now introduce a quantitive measurement of the boundedness of a continuous linear transformation.

Definition 9.5. Let $T : X \to Y$ be a bounded linear transformation. The infimum of all constants C such that

$$\|Tx\| \leq C\|x\|$$

for all $x \in X$ is the *operator norm* of T, denoted by $\|T\|_{X \to Y}$.

The name *operator* norm comes from classical analysis, in which an object that takes a function and transforms it was called an operator of functions. Since the spaces we consider in functional analysis are, by and large, abstractions of function spaces, much of the classical terminology remains. In fact, we shall often refer to linear transformations as *linear operators* from now on.

There are a number of different ways of computing the operator norm.

Remark 9.6. If $T : X \to Y$ is a bounded linear operator, then

$$\|T\|_{X\to Y} = \sup_{x\neq 0} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X=1} \|Tx\|_Y.$$

The name operator *norm* suggests that the collection of bounded linear operators should form a normed linear space, and this is indeed the case.

Remark 9.7. The space $\mathscr{B}(X, Y)$ of bounded linear operators from X to Y is a normed linear space with the operator norm.

Remark 9.8. If $T_n \to T$ in $\mathscr{B}(X, Y)$ and $x_n \to x$ in X, then $T_n x_n \to Tx$ in Y.

We remark that the notion of convergence in $\mathscr{B}(X, Y)$ is precisely the notion of convergence on the unit sphere in X. Specifically, we have the following result:

Remark 9.9. $\|T_n \to T\|_{X\to Y} \to 0$ if and only if

$$\sup_{\|x\|_X=1} \|T_n x - Tx\|_Y \to 0.$$

Theorem 9.10. *If Y is a Banach space, then so is $\mathscr{B}(X, Y)$.*

Proof. Let $(T_n)_{n=1}^\infty$ be a Cauchy sequence in $\mathscr{B}(X, Y)$. For each $x \in X$, the sequence $(T_n x)_{n=1}^\infty$ is Cauchy in Y, as

$$\|T_n x - T_m x\| \leq \|T_n - T_m\|\|x\|.$$

By the completeness of Y, we can find a limit Tx of the sequence $(T_n x)_{n=1}^\infty$. T is clearly linear. We also observe that

$$\|Tx\| \leq \|Tx - T_n x\| + \|T_n x\| \leq \|Tx - T_n x\| + \|T_n\|\|x\|.$$

Since $(T_n)_{n=1}^\infty$ is Cauchy, $M = \sup_n \|T_n\|$ is finite, and so

$$\|Tx\| \leq \|Tx - T_n x\| + M\|x\|.$$

We now send $n \to \infty$ to conclude that

$$\|Tx\| \leq M\|x\|,$$

and so T is bounded. Therefore, $T \in \mathscr{B}(X, Y)$.

It remains to show that $\|T_n - T\| \to 0$. To this end, we fix $\varepsilon > 0$ and find $N \in \mathbb{N}$ such that $\|T_n - T_m\| < \varepsilon$ for all $m > n > N$. Then $\|T_n x - T_m x\| < \varepsilon\|x\|$ for all $x \in X$, and sending $m \to \infty$ yields

$$\|T_n x - Tx\| < \varepsilon\|x\|.$$

This implies that $\|T_n - T\| < \varepsilon$ for all $n > N$, and so $T_n \to T$ in $\mathscr{B}(X, Y)$, as was to be shown. □

Let us now consider a few important special cases of the space of bounded linear operators. If the target space Y is the ground field $\mathbb{F}$, we obtain the space of *linear functionals* that are continuous:

Definition 9.11. The *(topological) dual space* X^* of the normed linear space X is the space $\mathscr{B}(X,\mathbb{F})$ of bounded linear functionals.

The name *topological* dual space contrasts X^* with a related construct, the *algebraic dual space* X'. X' is defined to be the collection of all linear functionals on X, continuous or not (see Definition 6.4). We remark that X' always has a plenty of elements: just take a basis of X' (which requires the Axiom of Choice) and define a linear functional by assigning a scalar to each basis element. The topological dual, however, is not always interesting: for some spaces, the topogical duals contain nothing but the zero functional, which sends every vector to zero. We shall have an occasion to say more about this later, when we discuss the Hahn-Banach theorem in §11 of this chapter.

In any case, we can apply Theorem 9.10 to $X^* = \mathscr{B}(X,\mathbb{F})$ even if X^* is trivial. And we do so at once:

Corollary 9.12. *X^* is a Banach space.* □

10. Completeness: Applications of the Baire Category Theorem

With the foresight that the topological dual space X^* is nontrivial in plenty of cases, we now exploit the *completness* of Banach spaces to establish several powerful results. These results will be derived as corollaries of the *Baire category theorem*, which, in turn, deals with a characterization of the notion of *smallness* in topological spaces.

10.1. **Nowhere Dense Sets and the Baire Category Theorem.** We begin by singling out an "obviously negligible" class of sets in topological spaces.

Definition 10.1. A subset A of a topological space X is *nowhere dense* in X if $(\overline{A})^\circ = \varnothing$.

Note that a finite union of nowhere dense sets is nowhere dense. Extending this idea, we can define a notion of smallness in topological spaces as follows:

Definition 10.2. A subset A of a topological space X is *of the first category*, or *meager*, if it is a countable union of nowhere dense sets. A is *of the second category* if A is not meager.

It is instructive to compare the notion of sets of category with the sets of measure zero. The next theorem shows that complete metric spaces cannot be too small.

Theorem 10.3 (Baire category theorem)**.** *A complete metric space is of the second category.*

Proof. We shall need the following lemma, which shows that open, dense subsets of a complete space cannot be too small:

Lemma 10.4. *If $\{U_n\}_{n=1}^\infty$ is a sequence of open, dense subsets of a complete metric space X, then $\bigcap_n U_n$ is also dense.*

Proof of lemma. Let V be an open subset of X. We shall show that the intersection of V and $\bigcap_n U_n$ is nontrivial. To this end, we first note that $U_1 \cap V$ contains a ball $B_{r_1}(x_1)$ centered at a point $x_1 \in U_1 \cap V$. We now define $B_{r_n}(x_n)$ inductively for $n > 1$ as follows: noting that $U_n \cap B_{r_{n-1}}(x_{n-1})$ has nonempty interior, we can find a ball $B_{r_n}(x_n)$ such that $B_{r_n}[x_n] \subseteq U_n \cap B_{r_{n-1}}(x_{n-1})$.

Let us relabel $(B_{r_n}(x_n))_{n=1}^\infty$ if necessary to have r_n tend to zero as $n \to \infty$. Then $(x_n)_{n=1}^\infty$ is Cauchy in the complete metric space X, whence $x_n \to x$ for some $x \in X$. By the construction,

$$x \in U_n \cap B_{r_1}(x_1) \subseteq U_n \cap V$$

for all n, which implies the desired conclusion. □

We now return to the task at hand. Suppose that $(E_n)_{n=1}^\infty$ is a sequence of sets such that $(\overline{E_n})^\circ = \varnothing$. For each $n \in \mathbb{N}$, the complement of $\overline{E_n}$ is open. In fact, it is dense, as

$$\overline{(\overline{E_n})^c} = [(\overline{E_n})^\circ]^c = X.$$

By the above lemma, we have

$$\bigcap_{n=1}^\infty (\overline{E_n})^c \neq \varnothing,$$

which implies that

$$\bigcup_{n=1}^\infty \overline{E_n} \neq X.$$

Therefore,

$$\bigcup_{n=1}^\infty E_n \neq X,$$

and the proof is complete. □

10.2. **The Open Mapping Theorem.** We now derive three powerful corollaries of the Baire category theorem. The first in line is the *open mapping theorem*, which provides a sufficient condition for a linear operator to be open. We first recall the definition of openness:

Definition 10.5. A function $f : X \to Y$ between topological spaces X and Y is *open* if the image of each open subset of X is open in Y.

The open mapping theorem asserts that surjectivity is enough if we are dealing with bounded linear operators between normed linear spaces.

Theorem 10.6 (Banach-Schauder, the open mapping theorem)**.** *If $T : X \to Y$ is a surjective bounded linear operator between Banach spaces X and Y, then T is open.*

Proof. We simplify our task by making the following observation:

Lemma 10.7. *Let $T : X \to Y$ be a linear operator between normed linear spaces X and Y. Then T is open if and only if $T(B_1(0))$ has nonempty interior.*

Proof of lemma. If T is open, then $T(B_1(0))$ is a nonempty open set, whence $T(B_1(0))$ must have a nonempty interior. Conversely, if $T(B_1(0))$ has a nonempty interior, then Proposition 8.4 implies that $T(B_r(x))$ must be open for each $r > 0$ and every $x \in X$. We now let E be an arbitrary open set in X. For each $x \in E$, we can find a real number $r > 0$ such that $B_r(x) \subseteq E$, whence $T(B_r(x))$ is an open neighborhood of Tx contained in $T(E)$. It follows that $T(E)$ is open. □

In light of this, we shall show that $\overline{T(B_1(0))}$ contains a ball, and generalize this result to show that $\overline{T(B_{1/2^n}(0))}$ contains a ball regardless of $n \in \mathbb{N}$. A strengthening of this argument will allow us to remove the closure, and the desired result will follow from Lemma 10.7.

We first prove that $\overline{T(B_1(0))}$ has nonempty interior[2]. Since $X = \bigcup_n B_n(0)$, we can write $Y = \bigcup T(B_n(0))$ by the surjectivity of T. The linearity of T allows us to rewrite the latter identity as

$$Y = \bigcup_{n=1}^{\infty} nT(B_1(0)),$$

whence the Baire category theorem (Theorem 10.3) applied to Y implies that at least one of the $nT(B_1(0))$ must fail to be nowhere dense. Since $nT(B_1(0))$ is homeomorphic to $T(B_1(0))$, it follows that $T(B_1(0))$ fails to be nowhere dense. In other words, $\overline{T(B_1(0))}$ has nonempty interior. We can now pick a point y_0 in the interior of $\overline{T(B_1(0))}$ and find a ball $B_r(y_0)$ contained in the interior of $\overline{T(B_1(0))}$.

Let us now show that $\overline{T(B_1(0))}$ contains a ball *centered at the origin*. Since $y_0 \in \overline{T(B_1(0))}$, we can find a point $y_1 \in T(B_1(0))$ and $x_1 \in B_1(0)$ such that $y_1 = Tx_1$ and $\|y_1 - y_0\| < r/2$. This, in particular, implies that the ball $B_{r/2}(y_1)$ is contained in $B_r(y_0)$, which, in turn, is contained in $\overline{T(B_1(0))}$.

Now, if $\|y\| < r/2$, then $y + y_1 \in B_{r/2}(y_1) \subseteq \overline{T(B_1(0))}$, whence

$$y = (y + y_1) - y_1 = (y + y_1) - Tx_1$$

is in $\overline{T(B_2(0))}$. Therefore, $\overline{T(B_2(0))}$ contains the ball $B_{r/2}(0)$, and a simple scaling argument shows that

$$B_{r/2^{n+2}}(0) \subseteq \overline{T(B_{1/2^n}(0))} \tag{10.8}$$

for each $n \in \mathbb{N}$.

Finally, we show that

$$T(B_1(0)) \supseteq B_{r/8}(0). \tag{10.9}$$

To this end, we fix a point $z \in B_{r/8}(0)$. (10.8) with $n = 1$ allows us to pick a point $p_1 \in B_{1/2}(0)$ such that $z - Tp_1 \in B_{r/16}(0)$. Similarly, we pick $p_k \in B_{1/2^k}(0)$ inductively such that

$$\left(z - \sum_{i=1}^{k} Tp_i \right) \in B_{r/2^{k+3}}(0),$$

applying (10.8) with $n = k$. The end result is a sequence $(p_k)_{k=1}^{\infty}$ in the Banach space X with

$$\sum_{k=1}^{\infty} \|p_k\| < \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

whence the following lemma implies that $\sum p_k$ must converge to a point $p \in B_1(0)$:

Lemma 10.10. *If $\sum \|x_n\| < \infty$, then $\sum x_n$ convergences.*

[2]This is *not* the same thing as $T(B_1(0))$ having nonempty interior, which is the desired result.

Proof of lemma. If $\sum_{n=1}^{\infty} \|x_n\| < \infty$, then the tail $\sum_{n=N}^{\infty} \|x_n\|$ can be made as small as desired, whence the partial sums $\left(\sum_{n=1}^{N} x_n\right)_{N=1}^{\infty}$ form a Cauchy sequence:

$$\left\| \sum_{n=1}^{M} x_n - \sum_{n=1}^{N} x_n \right\| = \left\| \sum_{n=N}^{M} x_n \right\| \leq \sum_{n=N}^{M} \|x_n\| \to 0 \text{ as } N, M \to \infty.$$

It follows that the series $\sum x_n$ converges. □

By continuity of T, we have

$$\|z - Tp\| = \lim_{k\to\infty} \left\| \left(z - \sum_{i=1}^{k} Tp_i \right) \right\| = 0,$$

and so $z \in T(B_1(0))$, thus establishing (10.9). This completes the proof. □

Remark 10.11. The converse of Lemma 10.10 also holds. Indeed, a normed linear space X is complete if and only if $\sum \|x_n\| < \infty$ implies the convergence of $\sum x_n$.

An immediate corollary of the open mapping theorem is that it is not very difficult to upgrade a linear isomorphism between two Banach spaces to an isomorphism of topological vector spaces.

Corollary 10.12. *A bijective bounded linear operator between Banach spaces is an isomorphism of topological vector spaces.*

Proof. It suffices to show that the inverse is continuous, which is equivalent to the openness of the operator in question. □

Since there is always a *bounded* linear isomorphism on a vector space onto itself, it follows that all complete norms on a single vector space must produce the same topology.

Corollary 10.13. *Let $\|\cdot\|$ and $\|\cdot\|'$ be two complete norms on a vector space X. If there exists a constant K_2 such that $\|\cdot\|' \leq K_2\|\cdot\|$, then there exists a constant K_1 such that $K_1\|\cdot\| \leq \|\cdot\|'$, rendering two norms equivalent.*

Proof. The identity operator I from $(X, \|\cdot\|)$ to $(X, \|\cdot\|')$ is a bijective bounded linear operator, hence an isomorphism of topological vector spaces by the previous corollary. In particular, it is a homeomorphism, which is to say that the two norms are equivalent. □

10.3. **Closed Graph Theorem.** Let us move onto the second result, the *closed graph theorem*, which concerns the *graph*

$$\Gamma(T) = \{(x, Tx) \in X \times Y : x \in X\}$$

of a linear operator $T : X \to Y$. We note that $\Gamma(T)$ is a linear subspace of the vector space $X \times Y$.

Remark 10.14. $X \times Y$ with componentwise scalar multiplication and vector addition is a vector space, isomorphic to the (external) direct sum $X \oplus Y$.

In particular, if X and Y are normed linear spaces, then $\Gamma(T)$ inherits a norm from the direct sum $X \times Y$, as per the following exercise.

Remark 10.15. If X and Y are normed linear spaces, then $\|\cdot\|_{X\times Y}$ given by

$$\|(x,y)\|_{X\times Y} = \|x\|_X + \|y\|_Y$$

is a norm on $X \times Y$ that generates the product topology. We refer the reader to Section 14 for a definition of the product topology.

We also observe that $\Gamma(T)$ is closed in $X \times Y$ if and only if $x_n \to x$ in X and $T(x_n) \to y$ in Y implies that $y = Tx$. While this is a consequence of the continuity of T, it is *a priori* a weaker condition than continuity.

Remark 10.16 ([Tao10], Example 1.7.18). The operator on the space c_0 of sequences vanishing at infinity that maps (a_n) to (na_n) is unbounded, but is graph is closed.

The next theorem shows that it is equivalent to continuity, provided that the spaces in question are complete.

Theorem 10.17 (Closed graph theorem). *A linear operator $T : X \to Y$ between Banach spaces X and Y is bounded if and only if $\Gamma(T)$ is closed.*

Proof. We have already established the "only if" part of the theorem. To show the "if" part, we assume that $\Gamma(T)$ is closed. Since X and Y are complete, $X \times Y$ is also complete, whence the closed linear subspace $\Gamma(T)$ of $X\times Y$ is complete as well.

We now consider the projection map $\pi : \Gamma(T) \to X$ given by $\pi(x, Tx) = x$ for each $x \in X$. π is a bijective bounded linear operator, and so Corollary 10.12 implies that π is an isomorphism of topological vector spaces. In particular, π^{-1} is a bounded linear operator, hence we can find a positive constant C such that

$$\|x\| + \|Tx\| \le C\|x\| \tag{10.18}$$

for all $x \in X$.

Now, excluding the trivial case that T is the zero operator, we see that there is at least one x such that $\|Tx\| > 0$. Therefore, the constant C in (10.18) is strictly larger than 1, so we can rewrite (10.18) as

$$\|Tx\| \le (C-1)\|x\|.$$

It follows that T is bounded, as was to be shown. □

Remark 10.19 ([Tao10], Theorem 1.7.19). A linear operator $T : X \to Y$ between Banach spaces X and Y is bounded if and only if there exists a Hausdorff topology $\mathscr{T}$ on Y that contains strictly fewer open sets than the norm topology on Y such that $T : X \to (Y, \mathscr{T})$ is still continuous. We refer the reader to Section 14 for a detailed discussion on *weak topologies*.

A nice application of the closed graph theorem is the following theorem of Alexandre Grothendieck on closed subspaces of Lebesgue spaces. See Chapter 4, Theorem 4.2 in [SS11] for a proof.

Theorem 10.20 (Grothendieck). *If (X, μ) is a finite measure space, $1 \le p < \infty$, E a closed subspace of $L^p(X,\mu)$, and $E \subseteq L^\infty(X,\mu)$, then E is finite-dimensional.*

10.4. **Uniform Boundedness Principle.** The third and final theorem provides a convenient method of upgrading pointwise bounds to uniform bounds.

Theorem 10.21 (Banach-Steinhaus, the uniform boundedness principle)**.** *Let* Λ *be a nonempty index set, and, for each* $\alpha \in \Lambda$*, we let* $T_\alpha : X \to Y$ *be a bouned linear operator between two normed linear spaces* X *and* Y*. We set*

$$E = \left\{ x \in X : \sup_{\alpha \in \Lambda} \|T_\alpha x\| < \infty \right\}.$$

(1) If E *is of the second category in* X*, then*

$$\sup_{\alpha \in \Lambda} \|T_\alpha\| < \infty.$$

(2) If $E = X$ *and* X *is a Banach space, then*

$$\sup_{\alpha \in \Lambda} \|T_\alpha\| < \infty.$$

Proof. (2) is a trivial consequence of the Baire category theorem (10.3), and so it suffices to establish (1). To this end, we let

$$E_n = \left\{ x \in X : \sup_{\alpha \in \Lambda} \|T_\alpha x\| \leq n \right\},$$

so that $E = \bigcup_n E_n$. Note that each E_n is closed, for

$$E_n = \bigcap_{\alpha \in \Lambda} \{x \in X : \|T_\alpha x\| \leq n\}.$$

E is of the second category, and so we can find some $n_0 \in \mathbb{N}$ such that E_{n_0} has nonempty interior. Therefore, we can find $x_0 \in X$ and $r > 0$ such that $B_r(x_0) \subseteq E_{n_0}$, whence $\|T_\alpha x\| \leq n_0$ for all $\alpha \in \Lambda$ whenever $\|x - x_0\| < r$. We now see that each $\|y\| < r$ satisfies the estimate

$$\|T_\alpha y\| \leq \|T_\alpha(y + x_0)\| + \|T(-x_0)\| = \|T_\alpha(y + x_0)\| + \|T(x_0)\| \leq 2n_0$$

for all $\alpha \in \Lambda$. Therefore, if $\|x\| \leq 1$, then we have

$$\|T_\alpha x\| = \|r^{-1} T_\alpha(rx)\| \leq \frac{2n_0}{r}$$

for all $\alpha \in \Lambda$, and so $\|T_\alpha\|$ is uniformly bounded. □

As an application of the uniform boundedness principle, we discuss the Fourier series. Recall that the *Fourier series* of a complex-valued L^1-function f on the circle group $\mathbb{T} \cong \mathbb{R}/\mathbb{Z}$ is given by

$$\sum_{n \in \mathbb{Z}} \hat{f}(n) e^{2\pi i n x}, \tag{10.22}$$

where

$$\hat{f}(n) = \int_0^1 f(t) e^{-2\pi i n t} \, dt$$

for each $n \in \mathbb{Z}$. Here we are considering f as a 1-periodic function on $\mathbb{R}$.

It is a classical result that (10.22) converges to f in the $L^2(\mathbb{T})$-norm if $f \in L^2(\mathbb{T})$. In other words,

$$\lim_{N \to \infty} \left\| f(x) - \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n x} \right\|_2 = 0.$$

This, of course, does not imply that the series converges pointwise to f. Investigating pointwise convergence of functions in a large space like L^1 or L^2 is difficult, so we opt to study a smaller space: the space $\mathscr{C}(\mathbb{T})$ of continuous complex-valued functions on $\mathbb{T}$. As it turns out, however, we can find, for each $x_0 \in \mathbb{T}$, a function $f \in \mathscr{C}(\mathbb{T})$, whose Fourier series diverges at x_0.

To see this, we assume without loss of generality that $x_0 = 0$. Observe that

$$\begin{aligned}(S_N f)(x) &= \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi i n x} \\ &= \int_{\mathbb{T}} f(t) \left(\sum_{n=-N}^{N} e^{2\pi i n (x-t)} \right) dt \\ &= \int_{\mathbb{T}} f(t) D_N(x-t)\, dt,\end{aligned}$$

where D_N is the *nth Dirichlet kernel*

$$D_N(x) = \sum_{n=-N}^{N} e^{2\pi i n x}.$$

Remark 10.23. It is not very hard to show that

$$D_N(x) = \frac{\sin \pi(2N+1)x}{\sin \pi x}.$$

Show also that there is a fixed constant $c > 0$ such that $\|D_N\|_1 \geq c \log N$ for all $N \in \mathbb{N}$.

We define a linear functional $l_{N,x} : \mathscr{C}(\mathbb{T}) \to \mathbb{C}$ by setting

$$l_{N,x}(f) = (S_N f)(x).$$

Observe that

$$|l_{N,0}(f)| = \left| \int f(t) D_N(t)\, dt \right| \leq \int |f(t)||D_N(t)|\, dt \leq \|f\|_\infty \|D_N\|_1.$$

This is, in fact, an equality:

Remark 10.24. Show that $\|l_{N,x}\| = \|D_N\|_1$ for all $N \in \mathbb{N}$ and $x \in \mathbb{T}$.

Combining the two exercises above, we can deduce the lower bonud $\|l_{N,0}\| \geq c \log N$, whence $\sup_N \|l_{N,0}\| = \infty$. Since $\mathscr{C}(\mathbb{T})$ is complete, it follows from the uniform boundedness principle (Theorem 10.21) that $(l_{N,0})_{N=1}^{\infty}$ is not bounded pointwise. In other words, we can find a function $f \in \mathscr{C}(\mathbb{T})$ such that $l_{N,0}(f) = (S_N f)(0)$ does not converge to $f(0)$.

Remark 10.25. The end result can be strengthened as follows: if A is a countable subset of $\mathbb{T}$, then there exists an $f \in \mathscr{C}(\mathbb{T})$ such that $(S_N f(x))_{N=1}^{\infty}$ fails to converge for all $x \in A$.

To see this, we can show first that if X is a Banach space, Y a normed linear space, and $(T_\alpha)_\alpha$ a collection of bounded linear maps from X to Y, then

$$M = \{x \in X : \sup_\alpha \|T_\alpha x\| = \infty\}$$

is either empty or a dense G_δ set in X.

Is it possible to find a larger set on which the Fourier series of a continuous function diverges? A result of Lennart Carleson states that the Fourier series of an L^2-function must converge almost everywhere to the function. This, in particular, implies that the Fourier series of a continuous function cannot diverge on too large of a set. See Volume 2, Chapter 7 of [MS13], Chapter 11 of [Gra10b], or Chapter 7 of [Thi06] for modern expositions on the theorem and the techniques used in the proof.

11. Duality: The Hahn-Banach theorems

Let us now recall from Corollary 9.12 that the (topological) dual space X^* of continuous linear functionals on a normed linear space X over $\mathbb{F}$ is a Banach space. The discussion preceding the corollary establishes that the algebraic dual space X' always has many elements, but it is not clear whether there exists a nontrivial element in X^*. The goal of this section is to show that X^* has *plenty* of elements in the following sense: for each nonzero element $x \in X$, we can find $l \in X^*$ such that $\|l\| = 1$ and $l(x) = \|x\|_X$. It then follows immediately that X^* *separates the points of* X, viz., every pair of distinct points $x_1, x_2 \in X$ admit $l \in X^*$ such that $l(x_1) \neq l(x_2)$.

Before we prove the result for the general case, we discuss some examples.

Example 11.1. If X is finite-dimensional, then $X^* = X'$, and so X^* is nontrivial.

Example 11.2. If $X = L^p(\Omega, \mu)$ with $1 \leq p \leq \infty$, then the linear functional $l_g : L^p(\Omega, \mu) \to \mathbb{C}$

$$l_g(f) = \int fg \, d\mu$$

is shown to be bounded via Hölder's inequality, provided that $g \in L^q(\Omega, \mu)$ with $1/p + 1/q = 1$:

$$\|l_g(f)\| \leq \|g\|_q \|f\|_p.$$

Therefore, $L^q(\Omega, \mu)$ can be isometrically embedded into $(L^p(\Omega, \mu))^*$, and $(L^p)^*$ is nontrivial. In fact, the Riesz representation theorem for Lebesgue spaces shows that the isometric embedding is an isomorphism if $1 < p < \infty$, but we will not prove this fact in this course. This result extends to $p = 1$ if (Ω, μ) is, for example, σ-finite.

Example 11.3. In general, $(L^\infty)^*$ is *not* L^1, even if the measure space is σ-finite. For example, if (Ω, μ) is $\mathbb{N}$ with the counting measure, then the space $\mathscr{C}_0(\mathbb{N})$ of sequences that converge to zero is a proper subspace of the space $l^\infty(\mathbb{N})$ of bounded sequences. We shall show that $(\mathscr{C}_0(\mathbb{N}))^* \cong l^1(\mathbb{N})$.

The above example shows that every $a = (a_n) \in l^1$ can be identified with a bounded linear functional on l^∞. Therefore, a can also be identified with a bounded linear functional on $\mathscr{C}_0$: indeed, if $a \in l^1$ and $x \in \mathscr{C}_0$, then

$$\left| \sum_n a_n x_n \right| \leq \|a\|_1 \|x\|_\infty$$

by Hölder's inequality. Therefore, there is an isometric embedding of l^1 into $\mathscr{C}_0$.

Conversely, we fix $l \in (\mathscr{C}_0)^*$. We shall show that l can be represented by an element of l^1. To this end, we define, for each $k \in \mathbb{N}$, a sequence $e^k = (e^k_n)_{n=1}^\infty$ by setting

$$e_n^k = \begin{cases} 1 & \text{if } k = n; \\ 0 & \text{if } k \neq n. \end{cases}$$

Then, for each $x = (x_k)_{k=1}^{\infty} \in \mathscr{C}_0$, we see that

$$l(x) = l\left(\sum_{k=1}^{\infty} x_k e^k\right) = \sum_{k=1}^{\infty} x_k l(e^k).$$

But then

$$|x_k l(e^k)| \leq |x_k||l(e^k)| \leq \|x\|_\infty |l(e^k)|,$$

and so

$$|l(x)| \leq \|x\|_\infty \|l(e^k)\|_1.$$

It follows that l^1 is isometrically isomorphic to $(\mathscr{C}_0)^*$. □

Let us now turn to the business of stating the theorem precisely. Here we work with a generalization of a norm.

Definition 11.4. A *seminorm* on a vector space X over $\mathbb{F}$ is a map $\rho : X \to [0, \infty)$ such that $\rho(\lambda x) = |\lambda|\rho(x)$ and $\rho(x+y) \leq \rho(x) + \rho(y)$ for all $\lambda \in \mathbb{F}$ and $x, y \in X$.

We remark that the L^p-norm is *a priori* a seminorm and only becomes a norm once the almost-everywhere equality formalism is introduced. See Section 15 for other examples of seminorms that arise naturally in analysis.

Theorem 11.5 (Complex Hahn-Banach)**.** *Let X be a vector space over $\mathbb{C}$. Let Y be a linear subspace of X, l a $\mathbb{C}$-linear functional on Y, and ρ a seminorm on X that dominates l on Y, viz., $|l(y)| \leq \rho(y)$ for all $y \in Y$. We can then find a $\mathbb{C}$-linear functional L on X that extends l and is dominated by ρ on X.*

The complex Hahn-Banach theorem will be derived from an analogous version of the theorem over $\mathbb{R}$. We now that we can relax the hypothesis in the real case to allow for sublinear functionals.

Definition 11.6. A *sublinear functional* on a vector space X is a map $\rho : X \to [0, \infty)$ such that $\rho(\lambda x) = \lambda\rho(x)$ and $\rho(x+y) \leq \rho(x) + \rho(y)$ for all $\lambda \geq 0$ and $x, y \in X$.

Theorem 11.7 (Real Hahn-Banach)**.** *Let X be a vector space over $\mathbb{R}$. Let Y be a linear subspace of X, l an $\mathbb{R}$-linear functional on Y, and ρ a sublinear functional on X that dominates l on Y, viz., $l(y) \leq \rho(y)$ for all $y \in Y$. We can then find an $\mathbb{R}$-linear functional L on X that extends l and is dominated by ρ on X.*

Proof of $\mathbb{R}$-HB $\Rightarrow$ $\mathbb{C}$-HB. We first recall that a $\mathbb{C}$-linear functional l on X can be written as the sum of two $\mathbb{R}$-linear functionals $l_1 = \operatorname{Re} l$ and $il_2 = i \operatorname{Im} l$. Simple computations show that $l_2(x) = -l_1(ix)$.

We now suppose that the $\mathbb{C}$-linear functional l on a subspace M of X is dominated by a seminorm ρ on X. This, in particular, implies that $l_1(x) \leq \rho(x)$ for all $x \in M$, whence the real Hahn-Banach theorem implies that we can extend l_1 to an $\mathbb{R}$-linear functional L_1 on X such that $L_1(x) \leq \rho(x)$ for all $x \in X$. Note that

$$-L_1(x) = L_1(-x) \leq \rho(-x) = \rho(x),$$

so that $|L_1(x)| \leq \rho(x)$ for all $x \in X$.

We now set $L(x) = L_1(x) - iL_1(ix)$, which is an extension of l onto X. Pick $\theta(x)$ such that
$$L(x) = e^{i\theta(x)}|L(x)|,$$
and observe that
$$|L(x)| = e^{-i\theta(x)}e^{i\theta(x)}|L(x)| = e^{-i\theta(x)}L(x) = L(e^{-i\theta(x)}x).$$
As $|L(x)| \geq 0$, we see that $|L(x)| = L_1(e^{-i\theta(x)}x) = |L_1(e^{-i\theta(x)}x)|$. It follows that
$$|L(x)| = |L_1(e^{-i\theta(x)}x)| \leq \rho(e^{-i\theta(x)}x) = |e^{-i\theta(x)}|\rho(x) = \rho(x),$$
as was to be shown. □

Proof of $\mathbb{R}$*-HB.* We shall first demonstrate how to extend a linear functional by one dimension and then use Zorn's lemma. We assume for now that M is a proper subspace of an $\mathbb{R}$-vector space X, and l a linear functional on M dominated by a sublinear functional ρ on X. We fix $x_0 \in X \smallsetminus M$ and extend l onto $M \oplus \mathbb{R}x_0$.

Let us construction an extension $L : M \oplus \mathbb{R}x_0 \to \mathbb{R}$ of l that is still dominated by ρ. We shall define $L(x_0)$ such that $L(x_0) \leq \rho(x_0)$, and extend onto $M \oplus \mathbb{R}x_0$ by linearity:
$$L(x + \lambda x_0) = l(x) + \lambda L(x_0).$$
L is clearly linear. To check that $L(x + \lambda x_0) \leq \rho(x + \lambda x_0)$, it suffices to show that
$$L(x + x_0) \leq \rho(x + x_0) \quad \text{and} \quad L(x - x_0) \leq \rho(x - x_0). \tag{11.8}$$
Indeed, (11.8) implies that
$$L(x + \lambda x_0) = \lambda\left[L(\lambda^{-1}x + x_0)\right] \leq \lambda\rho(\lambda^{-1}x + x_0) = \rho(x + \lambda x_0)$$
if $\lambda > 0$, and
$$L(x + \lambda x_0) = -\lambda\left[L(-\lambda^{-1}x - x_0)\right] \leq -\lambda\rho(-\lambda^{-1}x - x_0) = \rho(x + \lambda x_0)$$
if $\lambda < 0$.

We now note that (11.8) is equivalent to
$$L(x_1) - \rho(x_1 - x_0) \leq L(x_0) \leq \rho(x_2 + x_0) - L(x_2)$$
for all $x_1, x_2 \in M$. It thus suffices to define $L(x_0)$ to safisfy this inequality, which is possible if and only if
$$\sup_{x_1 \in M}\left[l(x_1) - \rho(x_1 - x_0)\right] \leq \inf_{x_2 \in M}\left[\rho(x_2 + x_0) - l(x_2)\right]. \tag{11.9}$$
To see this, we observe that
$$l(x_1) + l(x_2) = l(x_1 - x_0) + l(x_2 + x_0) \leq \rho(x_2 + x_0) + \rho(x_1 - x_0),$$
for all $x_1, x_2 \in M$, so that
$$l(x_1) - \rho(x_1 - x_0) \leq \rho(x_2 + x_0) - l(x_2).$$
(11.9) now follows by taking the supremum on the left-hand side and the infimum on the right-hand side. We have therefore successfully extended our linear functional by one dimension.

We now turn to the general case. We let $\mathcal{P}$ be a collection of pairs $(\tilde{M}, \tilde{l})$ consisting of a subspace $\tilde{M}$ of X that contains M and a linear functional $\tilde{l}$ on $\tilde{M}$ that extends l such that $\tilde{l}$ dominates l on M. We define a partial order on $\mathcal{P}$ by setting $(\tilde{M}_1, \tilde{l}_1) \prec (\tilde{M}_2, \tilde{l}_2)$ if and only if $\tilde{M}_1 \subseteq \tilde{M}_2$ and $\tilde{l}_2|_{\tilde{M}_1} = \tilde{l}_1$. If $\{(\tilde{M}_\alpha, \tilde{l}_\alpha)\}_\alpha$ is a chain in $\mathcal{P}$, then the upper bound of the chain is given by $(\tilde{M}, \tilde{l})$, where $\tilde{M} = \bigcup_\alpha \tilde{M}_\alpha$ and

$\tilde{l}$ is defined to agree with $\tilde{l}_\alpha$ on $\tilde{M}_\alpha$. Zorn's lemma therefore furnishes a maximal element (L, X) of $\mathcal{P}$. This L is unique and is the desired extension of l onto X. □

Corollary 11.10. *Let X be a normed linear space over $\mathbb{F}$. For each $x_0 \in X \smallsetminus \{0\}$, we can find $l \in X^*$ such that $\|l\| = 1$ and $l(x) = \|x_0\|$.*

Proof. We fix $x_0 \in X \smallsetminus \{0\}$ and let $Y = \mathbb{F}x_0$. Define a linear functional $l : Y \to \mathbb{F}$ by $l(\lambda x_0) = \lambda \|x_0\|$ for each $\lambda \in \mathbb{F}$, so that $l(x_0) = \|x_0\|$. Furthermore, each $y \in Y$ admits a scalar $\lambda \in \mathbb{F}$ such that $y = \lambda x_0$, and so

$$|l(y)| = |l(\lambda x_0)| = |\lambda||l(x_0)| = |\lambda|\|x_0\| = \|\lambda x_0\| = \|y\|.$$

We now define a seminorm ρ on X by setting $\rho(x) = \|x\|$, so that $|l(y)| = \rho(y)$ for each $y \in Y$. By the complex Hahn-Banach theorem (11.5), we can extend l to a linear functional $L : X \to \mathbb{C}$ that is dominated by ρ on X, whence $L \in X^*$. In particular, we have

$$|Lx| \leq \rho(x) = \|x\|$$

for all $X \in X$, so we have $\|L\| \leq 1$. It now suffices to note that $|L(x_0)| = |l(x_0)| = \|x_0\|$, whence $\|L\| = 1$. □

By the definition of boundedness, we have

$$|l(x)| \leq \|l\|_{X^*} \|x\|_X$$

for each $l \in X^*$ and every $x \in X$. Instead of considering l as a function that takes x, we now consider x as a function taking l: more precisely, we define a linear functional $\hat{x}$ on X^* by setting $\hat{x}(l) = l(x)$. Then we can rewrite the above inequality to see that

$$|\hat{x}(l)| \leq \|x\|_X \|l\|_{X^*},$$

whence $\|\hat{x}\|_{X^{**}} \leq \|x\|_X$. It follows that $\hat{x}$ is an element of the *double dual* $X^{**} = (X^*)^*$ of X. Furthermore, if $x \neq 0$, then Corollary 11.10 furnishes $l \in X^*$ such that $\|l\| = 1$ and

$$\hat{x}(l) = l(x) = \|x\|,$$

and so $\|\hat{x}\|_{X^{**}} \geq \|x\|_X$. We conclude that the map $x \mapsto \hat{x}$ is an isometric embedding from X into X^{**}.

Since Corollary 9.12 implies that X^{**} is always complete regardless of the completeness of X, the isometric embedding given above is not an isomorphism in general. Even when X is a Banach space, X^{**} is not always isometrically isomorphic to X: see the $\mathscr{C}_0$ example above. The cases in which X^{**} *is* isometrically isomorphic form an important subclass of the class of Banach spaces, and we give them a name.

Definition 11.11. A Banach space X is *reflexive* if the canonical embedding $x \mapsto \hat{x}$ into X^{**} is an isometric isomorphism[3].

We remark that the canonical embedding $x \mapsto \hat{x}$ provides a way to construct the *completion* of an incomplete normed linear space that we alluded to in Theroem 8.13. Indeed, if we let Y be the image of the canonical embedding $x \mapsto \hat{x}$, then $\overline{Y}$ is a complete subspace of X that includes an isometrically isomorphic copy Y of X.

[3]This isomorphism is *natural*, in the sense that the construction of the isomorphism does not require a choice of basis: see §17.3 for a discussion.

We now establish two restatements of the Hahn-Banach theorem, specialized for particular purposes. We shall consider some applications of each restatement as well.

Theorem 11.12 (Analytic Hahn-Banach)**.** *Let X be a normed linear space, M a subspace of X, and l a bounded linear functional on M. Then there exists a bounded extension L of l onto the whole space X such that $\|l\|_{M^*} = \|L\|_{X^*}$.*

Proof. Let $\rho(x) = \|l\|_{M^*}\|x\|$, so that ρ is a norm on X. Since $|l(x)| \leq \rho(x)$ for all $x \in M$, the complex Hahn-Banach theorem (Theorem 11.5) furnishes an extension L of l onto the whole space X that is still dominated by $\rho(x)$. Therefore, $\|L\|_{X^*} \leq \|l\|_{M^*}$. The reverse inequality is trivially established. □

Corollary 11.13 (Existence of nontrivial annihilators)**.** *Let X be a normed linear space over $\mathbb{F}$, M a subspace of X, and $x_0 \in X$ such that $d(x_0, M) > 0$. Then there exists $L \in X^*$ such that $L|_M = 0$, $L(x_0) = 1$, and $\|L\|_{X^*} = 1/d(x_0, M)$.*

See §19.2 for a discussion on the annihilators of subsets of a topological vector space.

Proof of corollary. Let $M_1 = M \oplus \mathbb{F}x_0$ and define a linear functional l on M_1 by setting $l(x + \lambda x_0) = \lambda$. Then $l = 0$ on M, and

$$\begin{aligned}
\|l\|_{M_1^*} &= \sup_{\substack{\lambda \neq 0 \\ x \in M}} \frac{|l(x + \lambda x_0)|}{\|x + \lambda x_0\|} \\
&= \sup_{\substack{\lambda \neq 0 \\ x \in M}} \frac{|\lambda|}{|\lambda| \|\lambda^{-1}x + x_0\|} \\
&= \frac{1}{\inf_{x,\lambda} \|\lambda^{-1}x + x_0\|} \\
&= \frac{1}{\inf_{u \in M} \|u + x_0\|} \\
&= \frac{1}{d(x_0, M)}.
\end{aligned}$$

The desired bounded linear functional L is now obtained as an extension of l via the analytic Hahn-Banach theorem (Theorem 11.12). □

Corollary 11.14. *If the dual space X^* of a Banach space X is separable, then so is X.*

Remark 11.15. $l^p(\mathbb{N})$ is a separable Banach space if and only if $1 \leq p < \infty$. In tandem with the above corollary, we conclude that $(l^\infty(\mathbb{N}))^* \neq l^1(\mathbb{N})$.

Proof of corollary. Let $\{l_n\}_{n \in \mathbb{N}}$ be a countable dense subset of X^*. We pick a sequence $(x_n)_{n=1}^\infty$ in X such that $\|x_n\| = 1$ and $|l_n(x_n)| \geq \frac{1}{2}\|l_n\|_{X^*}$ for all $n \in \mathbb{N}$. If we let M to be the span of $\{x_n\}_n$ and $\mathcal{L}$ the space of finite linear combinations of the elements of $\{x_n\}_n$ with rational coefficients, then $\mathcal{L}$ is a countable dense subset of M.

It therefore suffices to show that M is dense in X. We suppose for a contradiction that M is not dense in X, and pick $x \in X$ such that $d(x, M) > 0$. By Corollary

11.13, we can find a nonzero $L \in X^*$ that vanishes on M. Now, there exists a subsequence $(l_{n_k})_{k=1}^\infty$ of $(l_n)_{n=1}^\infty$ that converges to L in X^*. But

$$\|L - l_{n_k}\|_{X^*} \geq |(L - l_{n_k})(x_{n_k})| = |l_{n_k}(x_{n_k})| \geq \frac{1}{2}\|l_{n_k}\|_{X^*},$$

so that $\|l_{n_k}\|_{X*} \to 0$ as $n_k \to \infty$. This evidently contradicts the fact that L is nonzero, whence we must conclude that M is a dense subset of X. It follows that L is a dense subset of X, and X is separable. □

We conclude the section by introducing the notion of *adjoints*, which, in a sense, is a generalization of the complex conjugate operation on matrices.

Definition 11.16. The *(algebraic) adjoint* of a linear operator $T : X \to Y$ between two vector spaces X and Y is a linear operator $T' : Y' \to X'$ defined to be

$$T'y' = y'T$$

for each $y' \in Y'$.

If l is a linear functional on a vector space X, we often denote $l(x)$ by (l, x). Then the above definition states that $(T'y', x) = (y', Tx)$ for all $x \in X$ and $y' \in Y'$.

Definition 11.17. The *(topological) adjoint* of a bounded linear operator $T : X \to Y$ between two normed linear spaces X and Y is a linear operator $T^* : Y^* \to X^*$ defined to be

$$T^*y^* = y^*T$$

for all $y^* \in Y^*$.

We remark that T^*y^* is indeed an element of X^*, for it is a composition of two continuous maps. Furthermore, the adjoint operation preserves the operator norm.

Proposition 11.18. $\|T^*\| = \|T\|$

Proof. Observe that

$$\|T^*\| = \sup_{\|y^*\| \leq 1} \|T^*y^*\|_{X^*} = \sup_{\|y^*\| \leq 1} \sup_{\|x\| \leq 1} |T^*y^*(x)| = \sup_{\|y^*\| \leq 1} \sup_{\|x\| \leq 1} |y^*(Tx)|.$$

By the existence of nontrivial annihilators (Corollary 11.13), there exists a bounded linear functional y^* such that $\|y^*\| = 1$ and $y^*(Tx) = \|Tx\|$. Therefore,

$$\|T^*\| = \sup_{\|y^*\| \leq 1} \sup_{\|x\| \leq 1} |y^*(Tx)| \geq \sup_{\|x\| \leq 1} \|Tx\| = \|T\|.$$

Conversely,

$$\|T^*l\| = \|lT\| = \sup_{\|x\| \leq 1} \|lTx\| \leq \sup_{\|x\| \leq 1} \|l\|\|Tx\| = \|l\|\|T\| = \|T\|\|l\|$$

for all $l \in Y^*$, and so $\|T^*\| \leq \|T\|$. □

Also, bounded operators and their adjoints possess an injectivity-surjectivity duality relation.

Proposition 11.19. *Let X and Y be normed linear spaces. The adjoint T^* of $T \in \mathscr{B}(X, Y)$ is injective if and only if $\operatorname{im} T$ is dense in Y.*

Proof. Before we proceed, we observe that the injectivity of T^* is equivalent to the statement that $l_1 \circ T = l_2 \circ T$ implies $l_1 = l_2$ for all $l_1, l_2 \in Y^*$. This is a direct consequence of the definition of T^*.

($\Leftarrow$) We suppose that $\operatorname{im} T$ is dense in Y. If $l_1, l_2 \in Y^*$ satisfy $l_1 \circ T = l_2 \circ T$, then $l_1|_{\operatorname{im} T} = l_2|_{\operatorname{im} T}$, and so l_1 and l_2 are continuous functions that agree on a dense set. It follows that $l_1 = l_2$, and so T^* is injective.

($\Rightarrow$) Conversely, suppose that T^* is injective and assume for a contradiction that $\operatorname{im} T$ is not dense. Then $\overline{\operatorname{im} T}$ is a closed proper linear subspace of Y, whence the existence of nontrivial annihilators (Corollary 11.13) implies that there are $l_1, l_2 \in Y^*$ with $l_1|_{\operatorname{im} T} = l_2|_{\operatorname{im} T} = 0$ but $l_1 \neq l_2$. This contardicts the injectivity of T^*. □

The adjoint operation is a direct generalization of the conjugate transpose of square matrices: see Definition 16.45.

12. Convexity, Part 1: the Geometric Hahn-Banach Theorems

The next restatement of the Hahn-Banach theorem tells us why the proper functional-analytic context for the Hahn-Banach theorem is *convexity*. We begin with a few definitions.

Definition 12.1. Let X be a vector space over $\mathbb{R}$, and S a subset of X. A point $x_0 \in X$ is an *internal point* of S if, for each $u \in X$, there exists an $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $x_0 + tu \in S$.

We remark that all internal points of S are points in S. Furthermore, if X is a normed linear space, then every interior point of S is an internal point of S. The converse is true in $\mathbb{R}$, but even in $\mathbb{R}^2$ there are readily available counterexamples:

Example 12.2. Let

$$S = \{(r\cos\theta, r\sin\theta) : \theta - \pi \leq r \leq \pi - \theta,\ 0 \leq \theta < \pi\},$$

so that 0 is an internal point, but $0 \notin S^\circ$.

In general, if 0 is an internal point of a subset S of a real vector space X, then, for all $x \in X$, there exists a $\lambda_0 > 0$ such that $\lambda > \lambda_0$ implies $x \in \lambda S$. This motivates us to define the following:

Definition 12.3. The *Minkowski functional*, or the *gauge*, of a subset S of a real vector space X is given by the formula

$$\rho_S(x) = \inf\{\lambda > 0 : x \in \lambda S\}.$$

Let us note some basic properties of ρ_S:

Proposition 12.4. *The Minkowski functional ρ_S satisfies the following properties:*

(a) ρ_S is positive homogeneous, i.e., $\rho_S(\lambda x) = \lambda \rho_S(x)$ for all $\lambda \geq 0$ and $x \in S$.
(b) $\rho_S(x) \leq 1$ for all $x \in S$.
(c) If x is an internal point of S, then $\rho_S(x) < 1$.

Proof. (a) is trivial. To show (b), it suffices to note that $x \in 1S$ whenever $x \in S$. (c) follows from the fact that there exists a $\delta > 0$ such that $x + \delta x \in S$, so that $x \in (1+\delta)^{-1} S$. □

We now prove that the implication in (c) can be reversed if the underlying set is convex.

Proposition 12.5. *If K is a convex subset of a real vector space X containing 0 as an internal point, then x is an internal point of K if and only if $\rho_K(x) < 1$.*

Proof. It is enough to prove the "if" part. Suppose that $\rho_K(x) < 1$, and find $\delta > 0$ such that $x \in (1-\delta)K$. Since 0 is an internal point of K, each $u \in X$ admits an $\varepsilon > 0$ such that $|t| < \varepsilon$ implies $tu \in K$. We let $\varepsilon' = \varepsilon\delta$ and $t' = \delta t$. If $|t'| < \varepsilon'$, then $|t| < \varepsilon$, so that $tu \in K$. In particular, $t'u \in \delta K$. It now suffices to observe that $x \in (1-\delta)K$ and $t'u \in \delta K$ implies

$$x + t'u \in (1-\delta)K + \delta K = K,$$

as was to be shown. □

But what do Minkowski functionals have to do with the Hahn-Banach theorem? As it turns out, a Minkowski functional is a sublinear functional if the underlying set is convex.

Proposition 12.6. *If K is a convex subset of a real vector space X containing 0 as an internal point, then ρ_K is a sublinear functional on X.*

Proof. It suffices to check subadditivity of ρ_K. Let $x \in \lambda K$ and $y \in \mu K$, so that $\rho_K(x) \le \lambda$ and $\rho_K(y) \le \mu$. We now consider the representation

$$\frac{x+y}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\frac{x}{\lambda} + \frac{\mu}{\lambda+\mu}\frac{y}{\mu}.$$

Since x/λ and y/μ are in K, we see that $(x+y)/(\lambda+\mu)$ is in K, and so $x+y \in (\lambda+\mu)K$. It follows that $\rho_K(x+y) \le \lambda+\mu$. It now follows that

$$\rho_K(x+y) \le \inf_\lambda \lambda + \inf_\mu \mu = \rho_K(x) + \rho_K(y),$$

where the infima are taken over all values of λ and μ such that $\rho_K(x) \le \lambda$ and $\rho_K(y) \le \mu$. □

We are now in a position to state the first version of the geometric Hahn-Banach theorem.

Theorem 12.7 (Geometric Hahn-Banach for the algebraic dual)**.** *Let X be a real vector space and K_1 and K_2 two nonempty disjoint convex sets in X. If K_1 has an internal point, then there exists a nonzero $L \in X'$ such that $L(x) \le L(y)$ for all $x \in K_1$ and $y \in K_2$.*

In other words, l separates K_1 and K_2: indeed, there exists a real number α such that

$$\sup_{x \in K_1} l(x) \le \alpha \le \inf_{y \in K_2} l(y).$$

Proof. Let x_1 be an internal point of K_1 and x_2 a point of K_2. We note that $K_1 - K_2$ is convex, and that $x_1 - x_2$ is an internal point of $K_1 - K_2$. We let $K = K_1 - K_2 - x_0$, so that 0 is an internal point of K. Since K_1 and K_2 are disjoint, $0 \notin K_1 - K_2$, whence $-x_0 \notin K$. Proposition 12.5 now implies that $\rho_K(-x_0) \ge 1$.

Let $M = x_0\mathbb{R}$ and set $l(\lambda x_0) = -\lambda$ for each $\lambda \in \mathbb{R}$. l is a linear functional on M with $l(-x_0) = 1 \le \rho_K(-x_0)$. For each $\lambda \ge 0$, linearity of l and positive homogeneity of ρ_k imply that

$$l(-\lambda x_0) \le \rho_K(-\lambda x_0),$$

which, combined with the non-negativity of ρ_K, yields the estimate

$$l(\lambda x_0) = -\lambda \leq 0 \leq \rho_K(\lambda x_0).$$

Therefore, l is dominated by ρ_K, which, by Proposition 12.6, is sublinear. The real Hahn-Banach theorem (Theorem 11.7) now furnishes an extension $L \in X'$ of l dominated by ρ_K on X.

In particular, Proposition 12.4 (b) implies that the extension L is bounded above by 1 on K. We now fix arbitrary $y_1 \in K_1$ and $y_2 \in K_2$, respectively, and note that $y_1 - y_2 - x_0 \in K$. Therefore,

$$1 \geq L(y_1 - y_2 - x_0) = L(y_1) - L(y_2) + L(-x_0) = L(y_1) - L(y_2) + 1,$$

whence we have

$$L(y_1) \leq L(y_2).$$

We conclude that L is the desired linear functional. □

The second version of the geometric Hahn-Banach theorem can now be deduced as an easy consequence of the first one.

Corollary 12.8 (Geometric Hahn-Banach for the topological dual)**.** *Let X be a real normed linear space and K_1 and K_2 two nonempty disjoint convex sets in X. If K_1 has nonempty interior, then there exists a nonzero $L \in X^*$ such that $L(x) \leq L(y)$ for all $x \in K_1$ and $y \in K_2$.*

Proof. As in the above proof, we construct L on X. Note that $L(x) \leq \rho_K(x)$ and $-L(x) = L(-x) \leq \rho_K(-x)$, so that

$$-\rho_K(-x) \leq L(x) \leq \rho_K(x).$$

Note also that $K_1^\circ \neq \varnothing$ implies $0 \in K^\circ$, whence we can find a ball $B_{\varepsilon_0}(0) \subseteq K$.

We claim that ρ_K is continuous at 0. Indeed, for each $\varepsilon > 0$, we know that $B_\varepsilon(0) \subseteq \frac{\varepsilon}{\varepsilon_0} K$. It follows that $x_n \to 0$ implies $\rho_K(x_n) \to 0$, and the claim is established. In particular, L is continous at 0, hence continuous everywhere. Therefore, $L \in X^*$ □

For both cases, we remark that if all points of K_1 are internal, then all points of K are internal as well. Indeed, $K = \{x : \rho_K(x) < 1\}$, and so $L(x) < 1$ for all $x \in K$. It then follows that $L(y_1) < L(y_2)$ for all $y_1 \in K_1$ and $y_2 \in K_2$. We also remark that the same results hold for $\mathrm{Re}(L)$ if the underlying field is $\mathbb{C}$.

Remark 12.9. The nonempty interior assumption in Corollary 12.8 is there to guarantee the existence of internal points, which is crucial. Consider, for example, the subsets

$$E_\alpha = \{f \in \mathscr{C}([-1,1]) : f(0) = \alpha\}$$

of space $L^2([-1,1])$, which are dense and convex in L^2. Moreover, $E_\alpha \cap E_\beta = \varnothing$ whenever $\alpha \neq \beta$. Even so, E_α and E_β do not have any interior point, and there is no nonzero continuous linear functional l on L^2 for which

$$\sup_{f \in E_\alpha} l(f) \leq \inf_{f \in E_\beta} l(f).$$

Utilizing the additional geometric structure in Corollary 12.8, we now establish the following geometric result:

Theorem 12.10 (Hyperplane separation theorem). *Let X be a real normed linear space and A and B nonempty, convex, disjoint subsets of X. If A is compact and B is closed, then there exists an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in A} L(x) < \alpha < \inf_{y \in B} L(y).$$

In other words, the hyperplane $\{z : L(z) = \alpha\}$ separates A and B.

Proof. Consider the *Hausdorff metric*

$$d_H(A, B) = \inf_{\substack{x \in A \\ y \in B}} \|x - y\|.$$

Since $d_H(A, B) = \inf_{x \in A} d(x, B)$, the minimum is attained on the compact set A. Therefore, $d_H(A, B) > 0$.

Let $\varepsilon = d_H(A, B)/4$. We "beef up" A and B by setting

$$\tilde{A} = A + B_\varepsilon(0) \quad \text{and} \quad \tilde{B} = B + B_\varepsilon(0);$$

note that $\tilde{A}$ and $\tilde{B}$ are disjoint convex subsets of X with nonempty interior.The geometric Hahn-Banach theorem for the topological dual (Corollary 12.8) furnishes an $L \in X^*$ such that

$$\sup_{x \in \tilde{A}} L(x) \leq \alpha \leq \inf_{y \in \tilde{B}} L(y).$$

For each $x \in A$ and $u \in B_1(0)$, we see that $x + \varepsilon u \in \tilde{A}$. This, in turn, implies that

$$\sup_{u \in B_1(0)} L(x + \varepsilon u) = L(x) + \sup_{u \in B_1(0)} L(\varepsilon u) = L(x) + \varepsilon \|L\| \leq \alpha + \varepsilon \|L\|,$$

whence $L(x) < \alpha$ for each $x \in A$. Similarly, $L(y) > \alpha$ for all $y \in B$, and L is the desired linear functional. □

As a corollary, we deduce an alternate characterization of closed convex subsets of normed linear spaces. We shall need the a definition.

Definition 12.11. The *closed half-space* of a real vector space X with respect to $l \in X'$ at $\alpha \in \mathbb{R}$ is

$$H(l, \alpha) = \{x \in X : l(x) \leq \alpha\}.$$

We remark that $H(l, \alpha)$ is a closed subset of X if X is a normed linear space and l is bounded.

Corollary 12.12. *Every closed convex set K in a real normed linear space X can be written as the intersection of all closed half-spaces of X that contain K.*

Proof. Evidently, the intersection C of all closed half-spaces that contain K is a closed superset of K. To establish the reverse inclusion, we suppose for a contradiction that there exists $x_0 \in C \smallsetminus K$ and note that the singleton set $\{x_0\}$ is a compact, convex subset of X. Since K is a closed, convex subset of X disjoint from $\{x_0\}$, the hyperplane separation theorem (Theorem 12.10) applied to $\{x_0\}$ and K furnishes an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that

$$L(x_0) < \alpha < \inf_{x \in K} L(x).$$

In particular, $K \subseteq H(-L, -\alpha)$, and so $C \subseteq H(-L, -\alpha)$ by the minimality of C. Since $x_0 \in C$, it follows that $-L(x_0) \leq -\alpha$, or $L(x_0) \geq \alpha$, which contradicts the above inequality. □

13. Convexity, Part 2: the Krein-Milman Theorem

With the tools at hand, let us now carry out a detailed study of convex sets. We are, in particular, interested in the "generating sets" of convex sets. To make this notion precise, we make the following definition:

Definition 13.1. The *convex hull* of a subset E of a real vector space X is the smallest convex set $\operatorname{co}(E)$ that contains E, viz., the intersection of all convex sets containing E. Concretely, we can write $\operatorname{co}(E)$ as the collection

$$\left\{ \sum a_i v_i \,:\, a_i \in [0,1],\, \sum a_i = 1,\, v_i \in E \right\},$$

of convex combinations of elements of E

We can think of E as a generating set for the convex set $A = \operatorname{co}(E)$. To illustrate the utility of generating sets, we investigate the follwoing problem of maximizing convex functionals.

Proposition 13.2. *Let E be a nonempty, bounded subset of a real vector space X and $A = \overline{\operatorname{co}(E)}$. If f is a continuous convex functional on X, then*

$$\sup_{x \in A} f(x) = \sup_{x \in E} f(x).$$

Proof. Since f is continuous, the supremum of f over $A = \overline{\operatorname{co}(E)}$ agrees with the supremum over $\operatorname{co}(E)$. Observe that

$$\sup_{x \in \operatorname{co}(E)} f(x) \leq \sup_{\substack{x = \sum a_i v_i \\ \sum a_i = 1 \\ v_i \in E}} \sum a_i f(v_i) \leq \sum a_i \sup_{x \in E} f(x) = \sup_{x \in E} f(x),$$

and so $\sup_{x \in A} f(x) \leq \sup_{x \in E} f(x)$. The reverse inequality is trivial. □

Since $A = \operatorname{co}(E \cup F)$ for any subset $F \subseteq C$, it is quite possible that the generating set at hand contains many extraneous points. It is thus desirable to find a generating set E of a given convex set A that is "minimal" in the following sense: if D is a proper subset of E, then $\operatorname{co}(D) \neq A$. The set of vertices of a triangle is such a set—if any one of them is omitted, then the convex hull of the remaining two points is merely a line.

To make the notion of minimal generating sets precise, we fix two points x and y in a vector space X and consider the map $t \mapsto tx + (1-t)y$. The *closed line segment* $[x, y]$ and the *open line segment* $]x, y[$ from x to y are the images of $[0, 1]$ and $(0, 1)$ under this map, respectively.

Definition 13.3. A subset E of a subset A of a real vector space X is *extreme* if, for each $x, y \in A$, the nontriviality of the intersection $]x, y[\cap E$ implies that $x, y \in E$. A point $x \in A$ is an *extreme point* if $\{x\}$ is an extreme subset of A, and the collection of all such points of A is denoted by $\operatorname{ext}(A)$.

Note that extreme points are precisely the points that cannot be written as a convex combination of distinct points. As a conseqeunce, if E is a proper subset of $\operatorname{ext}(A)$, then $\operatorname{co}(E)$ cannot be A, for none of the points in $\operatorname{ext}(A)$ can be generated by taking a convex combination of points in $A \setminus \operatorname{ext}(A)$.

Example 13.4. The extreme points of a convex n-gon in $\mathbb{R}^2$ are precisely the vertices.

Example 13.5. Let $X = l^p(\mathbb{Z})$ and set $B = \{(a_n) \in X : \|(a_n)\|_p \leq 1\}$. If $p = 1$, then the extreme points of B are the sequence $(e_n^N)_{n=1}^{\infty}$ such that

$$|e_n^N| = \begin{cases} 1 & \text{if } n = N \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

If $1 < p < \infty$, then $\text{ext}(B) = \{(a_n) \in X : \|(a_n)\|_1 = 1\}$, where as $\text{ext}(B) = \{(a_n) \in X : |a_n| = 1 \text{ for all } n \in \mathbb{N}\}$ if $p = \infty$. This seems like a lot, but the unit ball in the subspace $_0$ of l^∞ has *no* extreme points.

Example 13.6. Let $X = L^p([0,1])$ and set $B = \{f \in X : \|f\|_p \leq 1\}$. If $p = 1$, then B has no extreme points, whereas $\text{ext}(B) = \{f \in X : \|f\|_p = 1\}$ if $1 < p < \infty$. If $p = \infty$, then

$$\text{ext}(B) = \{f \in X : |f(x)| = 1 \text{ for almost all } x \in [0,1]\}.$$

In particular, the unit ball in $\mathscr{C}([0,1])$ has only two extreme points: 1 and -1.

Example 13.7. Let Y be a compact metric space with the Borel σ-algebra. We define $\mathcal{M}(Y)$ to be the collection of all complex Borel measures on Y, and $\mathcal{A}(Y)$ to be the collection of all Borel probability measures on Y, viz., positive measures whose total measures are 1.

Given a continuous map $T : Y \to Y$, we say that $\mu \in \mathcal{A}(Y)$ is *T-invariant* if $\mu(T^{-1}B) = \mu(B)$ for every Borel set B in Y. The *Krylov-Bogoliubov theorem* guarantees that the collection $\mathcal{M}_T(Y)$ of T-invariant Borel probability measures is always nonempty. Recall now that μ is an *ergodic measure with respect to T* if $\mu(B \,\Delta\, T^{-1}(B)) = 0$ implies $\mu(B) = 0$ and $\mu(Y \setminus B) = 0$. A theorem of Oxtoby characterizes $\text{ext}_T(\mathcal{M}_T(Y))$ in terms of ergodic measures with respect to T.

While the set of extreme points is evidently minimal, they often do not serve as generating sets for the corresponding convex sets. Indeed, some very obviously convex sets, such as the L^1 unit ball, have no extreme points! The following theorem spells out useful criteria for extreme points serving as a generating set.

Theorem 13.8 (Krein-Milman). *If A is a nonempty, compact, convex subset of a normed linear space X, then $A = \overline{\text{co}(\text{ext}(A))}$.*

The rest of the section will be devoted to proving this theorem. To this end, we first estbalish a few basic properties of extreme sets and extreme points.

Proposition 13.9. *If F is an extreme subset of E and E an extreme subset of A, then F is an extreme subset of A.*

Proof. Let $x, y \in A$ and suppose that $y = tx + (1-t)y$ is in F for some $t \in (0,1)$. Since u is in the extreme subset E of A, the points x and y must be in E. F is an extreme subset of E, and so $u \in F$ implies that $x, y \in F$. □

Proposition 13.10. *Let f be a convex functional on a real vector space X. For each $A \subseteq X$, the preimage $E = (f|_A)^{-1}(\alpha)$ is an extreme subset of A, where $\alpha = \sup_{x \in A} f(x)$.*

Proof. If $x, y \in A$ and $tx + (1-t)y \in E$ for some $t \in (0,1)$, then

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

whence $tf(x) + (1-t)f(y) = \alpha$. Therefore, $f(x) = f(y) = \alpha$, and so $x, y \in E$. □

Let us now return to the task of proving the Krein-Milman theorem. To show that the extreme points serve as a generating set, we need to know that extreme points exist.

Lemma 13.11. *Every nonempty compact set K of a normed linear space X has an extreme point.*

Proof. We shall exhibit a minimal extreme subset of A and check that it is a singleton set. Let $\mathcal{E}_K$ be the collection of nonempty, compact, extreme subsets of K, ordered by reverse inclusion. $\mathcal{E}_K$ is nonempty, as K is in $\mathcal{E}_K$. If $(E_\alpha)_\alpha$ is a chain in $\mathcal{E}_K$, then we claim that $E = \bigcap_\alpha E_\alpha$ is the upper bound of the chain. It is clear that E is nonempty and compact. If $x, y \in K$ and $u = tx + (1-t)y$ is in E for some $t \in (0,1)$, then $u \in E_\alpha$, and so $x, y \in E_\alpha$, for each α. Therefore, E is an extreme subset of K, and thus an upper bound of $(E_\alpha)_\alpha$.

We apply Zorn's lemma on $\mathcal{E}_K$ to obtain the minimal extreme subset E_0 of K. If E_0 is not a singleton, then we can find two distinct points x_1 and x_2 in E_0. By the hyperplane separation theorem (Theorem 12.10), there exists an $l \in X^*$ such that $l(x_1) < l(x_2)$. We let $\beta = \sup_{x \in E_0} l(x)$ and set $E_0' = (l_{E_0})^{-1}(\beta)$. By the compactness of K, we can find $x_0 \in E_0'$, whence, in particular, E_0' is nonempty. E_0' is an extreme subset of E_0 as per Proposition 13.10, and thus an extreme subset of K by Proposition 13.9. Since $l(x_1) < l(x_2)$, it follows that $x_1 \in E_0 \smallsetminus E_0'$, which is now a nonempty extreme subset of K smaller than E_0, contradicting the minimality of E_0. We conclude that E_0 must be a singleton. □

The rest of the proof of the Krein-Milman theorem proceeds as follows:

Proof of the Krein-Milman theorem. Let A be a nonempty, compact, convex subset of a normed linear space X, and set $K = \overline{\text{co}(\text{ext}(A))}$. Since A is convex, $\text{co}(\text{ext}(A)) \subseteq A$, and the inclusion $K \subseteq A$ follows from the closedness of A.

To establish the reverse inclusion, we suppose for a contradiction that we can find an element $a \in A \smallsetminus K$. Applying the hyperplane separation theorem (Theorem 12.10) on K and $\{a\}$, we obtain $l \in X^*$ such that

$$\sup_{x \in K} l(x) < l(a). \tag{13.12}$$

Let $\alpha = \sup_{x \in A} l(x)$ and set $E = (l_A)^{-1}(\alpha)$. The compactness of A guarantees that E is nonempty. E is an extreme subset of K as per Proposition 13.10, and thus an extreme subset of A by Proposition 13.9. Furthermore, E is a closed subset of the compact set A, hence compact.

By Lemma 13.11, E has an extreme point e. Since $e \in E$, we see that $l(a) \leq l(e)$. On the other hand, $E \subseteq K$, and so (13.12) implies that $l(e) < l(a)$. This is evidently absurd, and we conclude that $A \smallsetminus K$ is empty, or that $A = K$. □

As a corollary, we now derive a useful strengthening of Proposition 13.2:

Corollary 13.13. *Let A be a nonempty, compact, convex subset of a normed linear space X and l a continuous convex functional on X. Then*

$$\sup_{x \in A} l(x) = \sup_{x \in \text{ext}(A)} l(x) \tag{13.14}$$

and there exists $a \in A$ such that $l(a) = \sup_{x \in A} l(x)$.

Proof. The Krein-Milman theorem (Theorem 13.8) shows that $A = \overline{\text{co}(\text{ext}(E))}$, and so Proposition 13.2 implies (13.14). Let $\alpha = \sup_{x \in A} l(x)$ and set $E = (l_A)^{-1}(\alpha)$. The compactness of A guarantees that E is a nonempty compact set, and Proposition 13.10 implies that E is an extreme subset of A. Lemma 13.11 furnishes an extreme point a of E, whence by Proposition 13.9 a is an extreme point of A. Since $a \in E$, we see that

$$l(a) = \sup_{x \in A} l(x),$$

as was to be shown. □

14. Weak and Weak-∗ Topologies

By now, the power and utility of convexity and compactness surely needs no further proof: we have seen, for example, the Krein-Milman theorem (Theorem 13.8), the geometric Hahn-Banach theorem (Theorem 12.8), and the hyperplane separation theorem (Theorem 12.10). Unfortunately, compact sets are quite hard to come by in a normed linear space. We have shown in Section 8 that the Heine-Borel property fails in normed linear spaces (Lemma 8.30), which, in particular, implies that closed balls are never compact. One conseqeunce is that a compact set in a normed linear space must have empty interior, for otherwise it contains a suitably small closed ball, implying that closed balls must be compact.

At the core of the problem is the over-abundance of open sets. The definition of compactness requires us to check every open cover, and a single failure to produce a finite subcover results in lack of compactness. A natural way to circumvent this problem is to reduce the number of open sets in a given space, thereby reducing the severity of this obstruction.

This method comes with a catch, however. We would like our functions and operators to be continuous whenever possible, and this property hinges on the abundance of open sets. Indeed, the inverse images of open sets would not be open if not too many sets were open to begin with.

14.1. The Weak Topology Paradigm and the Product Topology.

To find a happy medium, we consider the following paradigm: let X be a set, $\{Y_\alpha\}$ a collection of topological spaces, and, for each index α, a function $f_\alpha : X \to Y_\alpha$. Let us construct the *smallest* topology on X that turns all f_α continuous. Requiring the continuity of f_α amounts to defining $f_\alpha^{-1}(U_\alpha)$ be an open set in X whenever U_α is open in Y_α. This forces us to declare any union or finite intersection of these preimage. Collecting these open sets, we obtain a topology on X generated by the collection

$$\bigcup_\alpha \{f_\alpha^{-1}(U_\alpha) : U_\alpha \text{ open in } Y_\alpha\}_\alpha,$$

which we call the *weak topology with respect to the maps* f_α.

Let us consider an example. Recall that the *Cartesian product* of a collection $\{X_\alpha\}_{\alpha \in \mathcal{I}}$ of sets indexed by an index set $\mathcal{I}$ is defined to be the set $\prod_{\alpha \in \mathcal{I}} X_\alpha$ of functions $x : \mathcal{I} \to \bigcup_\alpha X_\alpha$ such that $x^\alpha = x(\alpha) \in X_\alpha$. Note that there is the *canonical projection map* $\pi_\alpha(f) = f(\alpha)$ for each $\alpha \in \mathcal{I}$. If each X_α is a topological space, then we can endow the *product topology* on the Cartesian product $\prod_\alpha X_\alpha$, which is the weak topology on $\prod_\alpha X_\alpha$ with respect to the canonical projection maps. Here we see that our aim of creating as many compact sets as possible is accomplished quite well:

Theorem 14.1 (Tychonoff)**.** *If $\{X_\alpha\}$ is a family of compact topological spaces, then the product topology on $\prod_\alpha X_\alpha$ turns $\prod_\alpha X_\alpha$ into a compact topological space.*

We sketch the proof, if only to introduce the notion of generalized sequences:

Sketch of proof. A *net*[4] in a topological space X is an X-valued function $\Phi : D \to X$ on a directed set D, viz., a partially ordered set in which every pair of elements α and β has a common maximum γ. A net Φ in X is *eventually in* a subset U of X if there exists an index $\alpha_0 \in D$ such that $\Phi(\alpha) \in U$ for all $\alpha \geq \alpha_0$. Φ *converges to* $x \in X$ if, for each open neighborhood U of x, the net Φ is eventually in U. Φ is *universal* if, for each subset A of X, Φ is eventually in either A or $X \smallsetminus A$.

Now, a topological space is compact if and only if every universal net therein converges. The image $\pi_\alpha \circ \Phi$ of a universal net Φ in $\prod_\alpha X_\alpha$ is universal in the compact space X_α and is thus convergent. Since a net Φ in $\prod_\alpha X_\alpha$ converges if and only if the components $\pi_\alpha \circ \Phi$ converge, the desired result follows. □

The component-wise convergence criterion for convergence of nets in a product space is worth reiterating. If $X_\alpha = X$ for all $\alpha \in \mathcal{I}$, then the Cartesian product $X^{\mathcal{I}} = \prod_{\alpha \in \mathcal{I}} X$ consists of X-valued functions $\varphi : \mathcal{I} \to X$ on $\mathcal{I}$. In this case, the component-wise convergence criterion states that a net (φ_β) of X-valued functions on $\mathcal{I}$ converges to $\varphi : \mathcal{I} \to X$ in the product topology of $X^{\mathcal{I}}$ if and only if

$$\pi_\alpha(\varphi_\beta) = \varphi_\beta(\alpha) \to \varphi(\alpha) = \pi_\alpha(\varphi)$$

for each $\alpha \in \mathcal{I}$.

Example 14.2. we consider $X = \mathbb{F}$ and $\mathcal{I} = [a, b]$. Here a net (φ_α) of $\mathbb{F}$-valued functions $\varphi_\alpha : [a, b] \to \mathbb{F}$ converges to φ in the product topology on $\mathbb{F}^{[a,b]}$ if and only if (f_α) converges pointwise to φ in the standard Euclidean topology on $\mathbb{F}$.

For this reason, the product topology is sometimes referred to as the *topology of pointwise convergence*. Using this idea, we can define a notion of boundedness on $X^{\mathcal{I}}$, provided that boundedness at each component is well-defined:

Definition 14.3. Let X be a metric space. A subset E in the product space $X^{\mathcal{I}}$ is *pointwise bounded* if

$$\pi_\alpha(E) = \{\varphi(\alpha) : \varphi \in E\}$$

is bounded for each $\alpha \in \mathcal{I}$.

It turns out that this notion of boundedness yields the Heine-Borel property for spaces of scalar-valued functions. We derive this as a quick corollary of Tychonoff's theorem.

Corollary 14.4 (Pointwise Arzelà-Ascoli theorem)**.** *A subset K of the product space $\mathbb{F}^{\mathcal{I}}$ with the product topology is compact if and only if K is closed and pointwise bounded.*

Proof. ($\Leftarrow$) Since the product of Hausdorff topological spaces is Hausdorff, $\mathbb{F}^{\mathcal{I}}$ is Hausdorff. Compact sets in a Hausdorff space are closed, and so K is closed.

[4]The details can be found in §19.3. For the remainder of this section, it suffices to think of nets as sequences with "uncountable index sets" as opposed to $\mathbb{N}$, retaining much of the properties that sequences in a topological space have. In analogy with sequences, we write $(x_\beta)_{\beta \in D}$ to denote a net with indices β in a directed set D.

Moreover, the continuous image $\pi_\alpha(K)$ of the compact set K is compact for each $\alpha \in \alpha \in \mathcal{I}$, hence bounded. It follows that K is pointwise bounded.

($\Rightarrow$) Let K be closed and pointwise bounded. $K_\alpha = \overline{\pi_\alpha(K)}$ is closed and bounded in $\mathbb{F}$, hence compact by the Heine-Borel theorem. By Tychonoff's theorem (Theorem 14.1), $\prod_{\alpha \in \mathcal{I}} K_\alpha$ is a compact subset of $\mathbb{F}^{\mathcal{I}}$. In particular, $\prod_\alpha K_\alpha$ contains the closed set K, whence K must be compact. □

14.2. **Weak-$*$ Topology.** Given a vector space X over $\mathbb{F}$, let us now consider the X-fold product

$$\mathbb{F}^X = \prod_{x \in X} \mathbb{F}$$

of the base field $\mathbb{F}$ with the product topology. $\mathbb{F}^X$ is precisely the set of $\mathbb{F}$-valued functions on X and includes the set X' of algebraic duals as its subset. If X is a topological vector space, then the topological dual X^* is also a subset of $\mathbb{F}^X$.

Definition 14.5. Let X be a topological vector space over $\mathbb{F}$. The *weak-$*$ topology* on X^* is the subspace topology with respect to the product topology on $\mathbb{F}^X$ and is denoted by $\sigma(X^*, X)$. We write $l_\beta \overset{*}{\rightharpoonup} l$ to denote the convergence of a net (l_β) to l in the weak-$*$ topology.

Once again, we see that our aim of obtaining many compact sets has been met:

Theorem 14.6 (Banach-Alaoglu). *Let X be a normed linear space. The norm unit ball*

$$B = \{l \in X^* : \|l\| \leq 1\}$$

is weak-$*$ compact. Consequently, every closed norm ball in X^* is weak-$*$ compact.

Proof. By the pointwise Arzelà-Ascoli theorem (Corollary 14.4), it suffices to check that B is weak-$*$ closed and pointwise bounded. If $l \in B$, then

$$|l(x)| \leq \|l\|_{X^*}\|x\|_X \leq \|x\|_X$$

for each $x \in X$, whence B is pointwise bounded.

To show that B is weak-$*$ closed, we take a net (l_β) in B that converges to some element $l \in \mathbb{F}^X$. l is evidently linear, and so it suffices to check that $\|l\|_{X^*} \leq 1$. For each $x \in X$, we have $l_\beta(x) \to l(x)$ by the component-wise convergence criterion, whence

$$|l(x)| = \lim_\beta |l_\beta(x)| \leq \lim_\beta \|x\| = \|x\|.$$

We conclude that $l \in B$. □

Remark 14.7. If X is separable, then an analogous result holds for sequential compactness: the norm unit ball of X^* is weak-$*$ sequentially compact. Indeed, if a normed linear space X is separable, then every weak-$*$ compact subset of X^* is metrizable (see [Rud91], Theorem 3.16). Since compactness and sequential compactness are equivalent on metrizable spaces, the weak-$*$ sequential compactness follows from Banach–Alaoglu. □

Let us now check that the weak topology paradigm that we have introduced in the beginning of the section has been adpoted in the construction of the weak-$*$ topology.

Proposition 14.8. *Let X be a topological vector space over $\mathbb{F}$. For each $x \in X$, the evaluation functional $\hat{x} : X^* \to \mathbb{F}$ by setting $\hat{x}(l) = l(x)$ for each $l \in X^*$ is a continuous linear functional on X^* with respect to $\sigma(X^*, X)$. Furthermore, $\sigma(X^*, X)$ is the weak topology generated by the collection of evaluation functionals.*

Proof. It suffices to mull over the fact that the evaluation functional $\hat{x}$ is precisely the canonical projection map $\pi_x : \mathbb{F}^X \to \mathbb{F}$ restricted to X^*. □

Using this characterization, we can give a simple description of the open sets in $\sigma(X^*, X)$.

Proposition 14.9. *Let X be a topological vector space. The sets*

$$V(x_1, \ldots, x_n; \varepsilon) = \{l \in X^* : |l(x_i) - l_0(x_i)| < \varepsilon \text{ for all } 1 \leq i \leq n\},$$

where $l_0 \in X^*$, $x_1, \ldots, x_n \in X$, and $\varepsilon > 0$, form a basis of neighborhoods of l_0 in the weak-$*$ topology $\sigma(X^*, X)$.

Proof. Let $y_i = l_0(x_i)$ and observe that

$$V(x_1, \ldots, x_n; \varepsilon) = \bigcap_{i=1}^{n} \widehat{x_i}^{-1}\left(B_\varepsilon(y_i)\right), \tag{14.10}$$

which is an open neighborhood of l_0 with respect to $\sigma(X^*, X)$ by Proposition 14.8. Conversely, the second half of Proposition 14.8 states that every open set in $\sigma(X^*, X)$ is generated by the preimages of open sets in $\mathbb{F}$ under the evaluation functionals, whence every neighborhood of l_0 can be written as a union of open sets of the form (14.10). The desired result now follows. □

Proposition 14.8 tells us that the evaluation functionals are elements of the dual space[5] $(X^\star)^*$ of the weak-$*$ dual $X^\star$ of X. If X is, for example, a normed linear space, then the Hahn-Banach theorem guarantees that these maps are distinct, so that $(X^\star)^*$ is nontrivial. We now show that every element of $(X^\star)^*$ is an evaluation functional, thereby proving that every normed linear space is reflexive[6] in the weak-$*$ sense.

Theorem 14.11. *Every normed linear space X is linearly isomorphic to the dual $(X^\star)^*$ of the weak-$*$ dual $X^\star$.*

We shall require two technical lemmas. The first provides a characterization of continuity of linear functionals using boundedness on a neighborhood.

Lemma 14.12. *A linear functional $l : Y \to \mathbb{F}$ on a topological vector sapce Y is continuous if and only if l is bounded on a neighborhood of 0.*

Proof of lemma. ($\Leftarrow$) If l is continuous, then $l^{-1}((-1, 1))$ is an open neighborhood of 0, on which l is bounded by 1.

($\Rightarrow$) It evidently suffices to prove that l is continuous at 0: indeed, a minor modification of the argument given in Proposition 9.4 would work. Suppose without loss of generality that l is bounded by 1 on a neighborhood U of 0. If (y_β) is a net in Y that converges to 0, then, for each $\varepsilon > 0$, we can find an index β_ε such that $\beta \geq \beta_\varepsilon$ implies $y_\beta \in \varepsilon U$, or $(1/\varepsilon) y_\beta \in U$. It now follows that $|l((1/\varepsilon) y_\beta)| \leq 1$,

[5] The use of $\star$ to denote the weak-$*$ dual is not standard and used only in this section for notational convenience.

[6] See Definition 11.11

whence $|l(y_\beta)| \le \varepsilon$ for all $\beta \ge \beta_\varepsilon$. Therefore, $l(y_\beta) \to 0$, and we conclude that l is continuous at 0. □

The second linear-algebraic lemma furnishes a necessary and sufficient condition for a linear functional to be a linear combination of other linear functionals.

Lemma 14.13. *If V is a vector space and $l_1, \dots, l_n, l$ are elements of the algebraic dual V' of V, then l is a linear combination of $l_1, \dots, l_n$ if and only if*

$$\bigcap_{i=1}^{n} \ker(l_i) \subseteq \ker l. \tag{14.14}$$

Proof of lemma. (⇒) If l is a linear combination of $l_1, \dots, l_n$, then $l_i(x) = 0$ for all i evidently implies $l(x) = 0$, which is precisely the statement of (14.14).

(⇐) We define $\pi : V \to \mathbb{F}^n$ by setting

$$\pi(x) = (l_1(x), \dots, l_n(x))$$

and consider the mapping $\pi(x) \overset{\Phi}{\mapsto} l(x)$ We first check that Φ is a well-defined linear functional on $\pi(V) \subseteq \mathbb{F}^n$. Indeed, if $\pi(x_1) = \pi(x_2)$, then $l_i(x_1) = l_i(x_2)$ for all $1 \le i \le n$, whence $x_1 - x_2 \in \bigcap \ker(l_i)$. (14.14) now implies that $x_1 - x_2 \in \ker l$, which, in particular, shows that

$$\Phi(\pi(x_1)) = l(x_1) = l(x_2) = \Phi(\pi(x_2)).$$

Now, as a linear functional on the finite-dimensional vector space $\pi(V)$, Φ admits a concrete representation:

$$\Phi(u_1, \dots, u_n) = \sum_{i=1}^{n} \lambda_i u_i.$$

It now follows that

$$l(x) = \Phi(\pi(x)) = \sum_{i=1}^{n} \lambda_i l_i(x),$$

as was to be shown. □

Proof of Theorem 14.11. The mapping $x \mapsto \hat{x}$ is evidently linear. If $\hat{x} = \hat{y}$, then $l(x - y) = 0$ for all $l \in X^*$. By the existence of nontrivial annihilators on a normed linear space (Proposition 11.13), it follows that $x - y = 0$, whence the mapping $x \mapsto \hat{x}$ is injective.

To show that the mapping $x \mapsto \hat{x}$ is surjective, we fix $\varphi \in (X^\star)^*$ and invoke Lemma 14.12 to find a neighborhood V of 0 on which φ is bounded by 1. By Proposition 14.9, we may assume without loss of generality that $V = V_{x_1,\dots,x_n;\varepsilon}$. Our goal is to write φ as a linear combination of $\widehat{x_1}, \dots, \widehat{x_n}$, which establishes the surjectivity of the mapping $x \mapsto \hat{x}$.

By Lemma 14.13, it suffices to check that

$$\bigcap_{i=1}^{n} \ker(\widehat{x_i}) \subseteq \ker \varphi.$$

Observe that if l is an element of the linear subspace $\bigcap \ker(\widehat{x_i})$ of $X^\star$, then $l(x_i) = 0$ for all $1 \le i \le n$, whence $\delta l \in V_{x_1,\dots,x_n;\varepsilon}$ for all $\delta > 0$. We see that $|\varphi(\delta l)| \le 1$, which implies that $|\varphi(l)| \le \delta^{-1}$ for all $\delta > 0$. It follows that $\varphi(l) = 0$, whence $l \in \ker \varphi$, as was to be shown. □

14.3. **Weak Topology.** We now employ the weak topology paradigm on the topological vector space X itself, rather than its dual.

Definition 14.15. The *weak topology* $\sigma(X, X^*)$ on a topological vector space X is the weak topology with respect to the elements of X^*, the continuous linear functionals. We write $x_\beta \rightharpoonup x$ to denote the convergence of a net (x_β) to x in the weak topology.

Recall that if X is a normed linear space, then the mapping $x \mapsto \hat{x}$ embeds X isometrically into X^{**}; let us write $\widehat{X}$ to denote the image. Certainly X^{**} is a subset of the product space $\mathbb{F}^{X^*}$, and so we may consider $\widehat{X}$ with the subspace topology inherited from the product topology on $\mathbb{F}^{X^*}$. Here the action of a typical canonical projection map looks like this:

$$\pi_l(\hat{x}) = \hat{x}(l) = l(x).$$

Therefore, the continuity of the projection maps π_l is equivalent to the continuity of the coressponding linear functionals, whence the subspace topology on $\widehat{X} \cong X$ inherited from the product topology on $\mathbb{F}^{X^*}$ is precisey the weak topology on X.

We now exhibit a simple collection of open sets that serves as a neighborhood basis in the weak topology, similar to that producted in Proposition 14.9.

Proposition 14.16. *Let X be a topological vector space. The sets*

$$V(l_1, \ldots, l_n; \varepsilon) = \{x \in X : |l_i(x) - l_i(x_0)| < \varepsilon \text{ for all } 1 \leq i \leq n\},$$

where $l_1, \ldots, l_n \in X^*$, $x_0 \in X$, and $\varepsilon > 0$, form a basis of neighborhoods of x_0 in the weak topology $\sigma(X, X^*)$.

Proof. Let $y_i = l_i(x_0)$ and observe that

$$V(l_1, \ldots, l_n; \varepsilon) = \bigcap_{i=1}^{n} l_i^{-1}\left(B_\varepsilon(y_i)\right), \tag{14.17}$$

which is an open neighborhood of x_0 with respect to $\sigma(X, X^*)$. Conversely, the definition of the weak topology on X states that every open set in $\sigma(X, X^*)$ is generated by the preimages of open sets in $\mathbb{F}$ under the bounded linear functionals, whence every neighborhood of x_0 can be written as a union of open sets of the form (14.17). The desired result now follows. □

Remark 14.18. Using the above characterization, we can show that a bounded set in an infinite-dimensional normed linear space can never be weak open. In particular, the weak topology on an infinite-dimensional space is always strictly weaker than the norm topology.

Remark 14.19. Nevertheless, it is possible for *strong convergence* and *weak convergence* to coincide on an infinite-dimensional space—for example, on $l^1(\mathbb{N})$. If $f_n \rightharpoonup f$ but $f_n \not\to f$, then we may assume without loss of generality that $f = 0$ and $\|f_n\| > 1$ and consider a subsequence $(f_{n_k})_{k=1}^\infty$ and an increasing sequence $(M_k)_{k=1}^\infty$ of indices such that

$$\|f_{n_k}\|_1 - \varepsilon < \sum_{n=M_{k-1}}^{M_k} |f_{n_k}(m)|$$

for all $k \in \mathbb{N}$. Letting

$$g(m) = \begin{cases} |f_{n_k}(m)|/f_{n_k}(m) & \text{if } M_{k-1}+1 \leq m \leq M_k \text{ for some } k; \\ 0 & \text{otherwise;} \end{cases}$$

we see that $|\sum_k f_{n_k}(m)g(m)| > 1 - 2\varepsilon$, which contradicts $f_n \rightharpoonup 0$.

Our next task is to characterize the dual of a normed linear space X with the weak topology. We say that a linear functional on X is *weakly continuous* if l is continuous with respect to the weak topology $\sigma(X, X^*)$. In contrast, a bounded linear functional on X is often said to be *strongly continuous*. The following proposition shows that this distinction is, in fact, unnecessary.

Proposition 14.20. *A linear functional l on X is weakly continuous if and only if it is strongly continuous.*

Proof. Let U be an arbitrary open set in $\mathbb{F}$. If l is weakly continuous, then $l^{-1}(U)$ is open in $\sigma(X, X^*)$, which is coarser than the norm topology on X. Therefore, $l^{-1}(U)$ is open in the norm topology of X, and l is strongly continuous. Conversely, if l is strongly continuous, then $l^{-1}(U)$ is an element of the generating set of $\sigma(X, X^*)$ and is thus open in $\sigma(X, X^*)$. It follows that l is weakly continuous. ☐

We remark that strong continuity always implies weak continuity, while weakly continuous *nonlinear* maps are often not strongly continuous.

Remark 14.21. [[Bre11], Exercise 4.20] We fix $p, q \in [1, \infty)$ and let $a : \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying the estimate

$$|a(t)| \leq C(|t|^{p/q} + 1)$$

for all $t \in \mathbb{R}$. Show that the operator $A : L^p(0,1) \to L^q(0,1)$ defined by the formula

$$(Af)(x) = a(f(x))$$

is strongly continuous but is never weakly continuous unless a is affine.

14.4. **Concluding Remarks.** For the sake of concreteness, we let X be a normed linear space. X can be endowed with either of the following two topologies.

(1) The *norm topology*, which gives rise to the strong convergence $x_\beta \to x$ of nets, i.e., $\lim_\beta \|x_\beta - x\| = 0$. This is equivalent to the **uniform** convergence of the corresponding nets $(\widehat{x_\beta})$ of evaluation functionals on the norm unit ball $B_1[0]$ in X^{**}, which can also be written as

$$\lim_\beta \sup_{\|l\| \leq 1} |\widehat{x_\beta}(l) - \hat{x}(l)| = \lim_\beta \sup_{\|l\| \leq 1} |l(x_\beta) - l(x)| = 0.$$

(2) The *weak topology* $\sigma(X, X^*)$, which gives rise to the weak convergence $x_\beta \rightharpoonup x$ of nets, i.e., $l(x_\beta) \rightharpoonup l(x)$ for all $l \in X^*$. This is equivalent to the **pointwise** convergence of the corresponding nets $(\widehat{x_\beta})$ of evaluation functionals in the double dual X^{**}.

X^*, on the other hand, allows for three possible topologies.

(1) The *norm topology*, which gives rise to the strong convergence $l_\beta \to l$ of nets of linear functionals. This is equivalent to the **uniform** convergence of the nets on the norm unit ball $B_1[0]$ in X.

(2) The *weak-∗ topology* $\sigma(X^*, X)$, which gives rise to the weak-∗ convergence $l_\beta \overset{*}{\rightharpoonup} l$ of nets of linear functionals. This is equivalent to the **pointwise** convergence of the nets (l_β) on X.
(3) The *weak topology* $\sigma(X^*, X^{**})$, which gives rise to the weak convergence $l_\beta \rightharpoonup l$ of nets of linear functionals, i.e., $\varphi(l_\beta) \to \varphi(l)$ for all $\varphi \in X^{**}$.

Let us now discuss two cases in which some of these topologies might be the same.

Proposition 14.22. *If X is finite-dimensional, then all of the topologies are the same.*

Proof. For notational convenience, we consider $\mathbb{R}^n$ with the Euclidean norm topology. We can do this without loss of generality, as all norms on a finite-dimensional vector space are equivalent (Theorem 8.18). The projection maps onto the kth coordinate $\pi_k(x) = x^k$ are continuous linear functionals on $\mathbb{R}^n$. Since the weak topology with respect to the projection maps $\pi_1, \ldots, \pi_n$ is already the Euclidean norm topology, the weak and norm topologies on $\mathbb{R}^n$ are the same. Similarly, the weak and norm topologies on $(\mathbb{R}^n)^* \cong \mathbb{R}^n$ are the same. The product-space definition of the weak-∗ topology implies that the weak-∗ topology on $(\mathbb{R}^n)^*$ is precisely the weak topology generated by the projection maps $\pi_1, \ldots, \pi_n$, which is the same as the weak and norm topologies on $(\mathbb{R}^n)^*$. □

Proposition 14.23. *If X is reflexive, then the weak-∗ topology $\sigma(X^*, X)$ and the weak topology $\sigma(X^*, X^{**})$ are the same.*

Proof. If X is reflexive, then every element of X^{**} is an evaluation functional. Therefore, the weak convergence $l_\beta \rightharpoonup l$ of nets of linear functionals in this case is equivalent to $\hat{x}(l_\beta) \to \hat{x}(l)$ for all $x \in X$. We can rewrite this as $l_\beta(x) \to l(x)$, which is precisely the pointwise-convergence criterion for the weak-∗ convergence $l_\beta \overset{*}{\rightharpoonup} l$ of nets of linear functionals. □

Much of our discussion in this section relied on net characterizations of continuity and compactness. These can be reduced to the usual sequential characterizations with the added assumption of metrizability. Unfortunately, metrizability turns out to be too strong of an assumption in general. Indeed:

Theorem 14.24. *The weak topology on X is metrizable if and only if X is finite-dimensional.*

Proof [Lec]. We shall require the following technical lemma.

Lemma 14.25. *The vector-space dimension of an infinite-dimensional Banach space is unctounable.*

Proof of Lemma. Let X be an infinite-dimensional Banach space and suppose for a contradiction that $\{v_n\}_{n\in\mathbb{N}}$ is a Hamel basis of X. For each $n \in \mathbb{N}$, we set $X_n = \operatorname{span}\{v_1, \ldots, v_n\}$. By Corollary 8.23, each X_n is complete, hence closed in X. Furthermore, X_n evidently cannot contain a norm ball, hence X_n has empty interior. It now follows that $X = \bigcup_n X_n$ is of the first category, which contradicts the Baire category theorem (Theorem 10.3). □

If X is finite-dimensional, then the weak topology is the same as the norm topology, which is evidently metrizable. Conversely, we suppose that the weak topology on X is induced by a metric d. We shall produce a countable spanning set of X^*. By Corollary 9.12, X^* is a Banach space, and so Lemma 14.25 forces X^* to be finite-dimensional. This, in particular, implies that X^* is linearly isomorphic to X^{**}. Since X embeds into X^{**}, it follows that X must be finite-dimensional as well.

For each $n \in \mathbb{N}$, we take the metric ball $B_{1/n}(0)$ of radius $1/n$ centered at 0. By Proposition 14.16, each n furnishes a finite collection $\mathcal{L}_n = \{l_{1,n}, \ldots, l_{k_n,n}\} \subseteq X^*$ and a positive scalar ε_n such that $\mathcal{V}_n = V_{l_{1,n},\ldots,l_{k_n,n};\varepsilon_n}$ is contained in $B_{1/n}(0)$. We shall show that $\mathcal{L} = \bigcup_n \mathcal{L}_n$ spans X^*.

To this end, we fix $l \in X^*$ and consider the weak neighborhood $V_{l;1}$ based at 0. There exists an $N \in \mathbb{N}$ such that $B_{1/N}(0) \subseteq V_{l;1}$, so that $\mathcal{V}_N \subseteq V_{l;1}$. We claim that l is a linear combination of the elements of $\mathcal{L}_N$. By Chapter Lemma 14.13, it suffices to show that

$$\bigcap_{i=1}^{k_N} \ker(l_{i,N}) \subseteq \ker l.$$

If $l_{i,N}(x) = 0$ for all $1 \leq i \leq k_N$, then $l_{i,N}(Mx) = 0$ for all $M > 0$, whence $Mx \in \mathcal{V}_n \subseteq V_{l;1}$. This, in particular, implies that $|l(Mx)| < 1$, or $|l(x)| < 1/M$ for all $M > 0$, whence $l(x) = 0$. The proof is now complete. □

Similarly, the weak-$*$ topology on X^* is metrizable if and only if X is finite-dimensional.

15. Locally Convex Spaces

The natural setting for many of the results developed so far is a topological vector space with sufficiently many open convex sets, which we define precisely as follows:

Definition 15.1. A topological vector space is said to be *locally convex* if it has a neighborhood basis consisting of convex sets.

We see at once that normed linear spaces are locally convex; this is a consequence of the triangle inequality. The following proposition provides a typical way of creating a locally convex space.

Proposition 15.2. *Let X be a vector space, $\{\rho_\alpha\}_{\alpha \in \mathcal{I}}$ a family of seminorms on X, x_0 a fixed point in X, and $\mathscr{B}(x_0)$ a collection of sets of the form*

$$V(\alpha_1, \ldots, \alpha_n; \varepsilon) = \{x \in X : \rho_{\alpha_i}(x - x_0) < \varepsilon \text{ for all } 1 \leq i \leq n\}.$$

Then $\mathscr{B}(x_0)$ is a basis of neighborhoods of x_0 consisting of convex sets, and the topology generated by

$$\bigcup_{x_0 \in X} \mathscr{B}(x_0) \tag{15.3}$$

turns X into a locally convex space.

In particular, if X is a normed linear space, then the weak topology $\sigma(X, X^*)$ turns X into a locally convex space (Proposition 14.16), and the weak-$*$ topology $\sigma(X^*, X)$ turns X^* into a locally convex space (Proposition 14.9).

Proof. Let X be endowed with the topology generated by (15.3). We observe that a net (x_β) in X converges to $x_0 \in X$ if and only if $\rho_\alpha(x_\beta - x_0) \to 0$ in for all α. Indeed, the seminorm convergence criterion shows that that each single-index neighborhood $V(\alpha;\varepsilon)$ admits an index β_α such that $\beta \geq \alpha$ implies $x_\beta \in V(\alpha;\varepsilon)$. We then see that x_β is in

$$V(\alpha_1,\ldots,\alpha_n;\varepsilon) = \bigcap_{i=1}^n V(\alpha_i;\varepsilon)$$

for all $\beta \geq \max\{\beta_1,\ldots,\beta_n\}$. Since $\mathscr{B}(x_0)$ is a basis of neighborhoods of x_0, it follows that $x_\beta \to x_0$. The converse is established analogously.

We now show that X is a topological vector space. If $((\lambda_\beta, x_\beta))_\beta$ be a net in $\mathbb{F} \times X$ that converges to (λ, x), then

$$\rho_\alpha(\lambda_\beta x_\beta - \lambda x) \leq \rho_\alpha(\lambda_\beta(x_\beta - x)) + \rho_\alpha((\lambda_\beta - \lambda)x) \to 0$$

for all α, whence the scalar multiplication operation is continuous. To show that vector addition is continuous, we fix a net (x_β, y_β) in $X \times X$ that converges to (x,y). We observe that

$$\rho_\alpha((x_\beta + y_\beta) - (x+y)) \leq \rho_\alpha(x_\beta - x) + \rho_\alpha(y_\beta - y) \to 0$$

for each α, whence $x_\beta + y_\beta \to x + y$. It follows that the vector addition operation is continuous, and X is a topological vector space.

It remains to show that X is a locally convex space. To this end, we observe that $V(\alpha;\varepsilon)$ based at x_0 is convex, for if $x, y \in V(\alpha;\varepsilon)$, then

$$\begin{aligned}
\rho_\alpha([(1-\lambda)x + \lambda y] - x_0) &= \rho_\alpha((1-\lambda)(x - x_0) + \lambda(y - x_0)) \\
&\leq (1-\lambda)\rho_\alpha(x - x_0) + \lambda\rho_\alpha(y - x_0) \\
&< \varepsilon,
\end{aligned}$$

whence $(1-\lambda)x + \lambda y \in V(\alpha;\varepsilon)$ for all $0 \leq \lambda \leq 1$. Since intersections of convex sets are convex, $\mathscr{B}(x_0)$ consists of convex sets, and the desired result follows. □

We now generalize the notion of bounded operators (Definition 9.3) via seminorms.

Proposition 15.4. *Let X and Y be locally convex spaces whose topologies are generated by $\{\rho_\alpha\}_\alpha$ and $\{\sigma_\beta\}_\beta$, respectively, à la Proposition 15.2. A linear operator $T : X \to Y$ is continuous if and only if each Y-index β admits a finite collection of X-indices $\{\alpha_1,\ldots,\alpha_n\}$ and a positive constant A such that*

$$\sigma_\beta(Tx) \leq A \sum_{i=1}^n \rho_{\alpha_i}(x) \tag{15.5}$$

for all $x \in X$.

Proof. ($\Leftarrow$) Let (x_γ) be a net in X that converges to $x \in X$. By Proposition 15.2, $\rho_\alpha(x_\gamma - x) \to 0$ for all α, whence $\sigma_\beta(Tx_\gamma - Tx) \to 0$ by (15.5). It follows that T is continuous.

($\Rightarrow$) Suppose that T is continuous, so that the preimage $T^{-1}(V(\beta;1))$ of

$$V(\beta;1) = \{y \in Y : \sigma_\beta(y) < 1\}$$

is an open neighborhood of 0. By Proposition 15.2, we can find a neighborhood $V(\alpha_1,\ldots,\alpha_n;\varepsilon)$ of 0 that is entirely contained in $T^{-1}(V(\beta;1))$.

We fix $x \in X$. If $\rho_{\alpha_i}(x) \neq 0$ for some $1 \leq i \leq n$, then we can define the vector

$$z = \left[\frac{\varepsilon}{2} \left(\sum_{i=1}^{n} \rho_{\alpha_i}(x) \right)^{-1} \right] x,$$

so that

$$\rho_{\alpha_i}(z) = \frac{\varepsilon}{2} \cdot \frac{\rho_{\alpha_i}(x)}{\sum \rho_{\alpha_i}(x)} < \varepsilon$$

for all $1 \leq i \leq n$. Therefore, $z \in V(\alpha_1, \ldots, \alpha_n; \varepsilon) \subseteq T^{-1}(V(\beta; 1))$, and we have

$$\frac{\varepsilon}{2} \frac{\sigma_\beta(Tx)}{\sum \rho_{\alpha_i}(x)} = \sigma_\beta(Tz) < 1,$$

whence it follows that

$$\sigma_\beta(Tx) \leq \frac{2}{\varepsilon} \sum_{i=1}^{n} \rho_{\alpha_i}(x),$$

as was to be shown.

On the other hand, if $\rho_{\alpha_i}(x) = 0$ for all $1 \leq i \leq n$, then $\rho_{\alpha_i}(Mx) = 0$ for all $M > 0$. This implies that $Mx \in V(\alpha_1, \ldots, \alpha_n; \varepsilon) \subseteq T^{-1}(V(\beta; 1))$ for all $M > 0$, which, in turn, shows that $\sigma_\beta(Tx) < 1/M$ for all $M > 0$. It follows that $\sigma_\beta(Tx) = 0$, and (15.5) is obtained trivially. □

As a corollary, we obtain a generalization of the notion of equivalence of norms (Definition 8.16).

Corollary 15.6. *Two collections of seminorms $\{\rho_\alpha\}_\alpha$ and $\{\sigma_\beta\}_\beta$ generate the same topology on a vector space X if and only if the following two conditions hold:*

(1) For each β, there exist indices $\alpha_1, \ldots, \alpha_n$ and a positive constant A such that

$$\sigma_\beta(Tx) \leq A \sum_{i=1}^{n} \rho_{\alpha_i}(x)$$

for all $x \in X$.

(2) For each α, there exist indices $\beta_1, \ldots, \beta_n$ and a positive constant B such that

$$\rho_\alpha(Tx) \leq B \sum_{i=1}^{n} \sigma_{\beta_i}(x)$$

for all $x \in X$.

Proof. For notational convenience, we write X_α and X_β to denote X with topologies generated by $\{\rho_\alpha\}$ and $\{\sigma_\beta\}$, respectively. If (1) and (2) hold, then Proposition 15.4 implies that the identity map $\mathrm{id} : X_\alpha \to X_\beta$ is a homeomorphism. Conversely, if $X_\alpha \cong X_\beta$, then $\mathrm{id} : X_\alpha \to X_\beta$ is a homeomorphism, whence (1) and (2) follow from Proposition 15.4. □

The generalizations we have carried out above suggest that much of the Banach-space theory could be carried over to an appropriate class of locally convex spaces. We now introduce a metrizable topological vector space that is complete with respect to the seminorms that generate the topology.

Definition 15.7. A *Fréchet space* is a topological vector space X whose topology is generated by a complete translation-invariant metric.

In what sense is a Fréchet space complete with respect to seminorms? The next three propositions detail a typical method of creating a Fréchet space.

Proposition 15.8. *The topology on a topological vector space X generated by a countable family of seminorms $\{\rho_n\}_{n\in\mathbb{N}}$ is metrizable by the following translation-invariant metric:*

$$d(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\rho_n(x-y)}{1+\rho_n(x-y)}.$$

Proof. For notational convenience, we write X_ρ and X_d to denote X with topologies generated by $\{\rho_n\}$ and d, respectively. Our goal is to show that X_ρ is homeomorphic to X_d. To do so, it suffices to show that an open set in X_ρ is open in X_d and vice versa.

($\Rightarrow$) By Proposition 15.2 and the translation invariance of the two topologies at hand, it suffices to show that

$$V(n_1,\ldots,n_k;\varepsilon) = \{x \in X : \rho_{n_i}(x) < \varepsilon \text{ for all } 1 \leq i \leq k\}$$

is open in X_d. We fix $x \in V(n_1,\ldots,n_k;\varepsilon)$. If $y \in X$ satisfies the lower bound $\rho_{n_i}(y) \geq \varepsilon$ for some $1 \leq i \leq n$, then

$$\begin{aligned}
d(x,y) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(x-y)}{1+\rho_{n_i}(x-y)} \\
\text{(triangle inequality for } \rho_{n_i}) &\geq \frac{1}{2^{n_i}} \frac{|\rho_{n_i}(x)-\rho_{n_i}(y)|}{1+\rho_{n_i}(x)+\rho_{n_i}(y)} \\
\text{(upper bound of } \rho_{n_i}(x)) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(y)-\rho_{n_i}(x)}{(1+\varepsilon)+\rho_{n_i}(y)} \\
\text{(removing dependence on } n_i) &\geq \frac{1}{2^{n_i}} \frac{\rho_{n_i}(y)-\max\limits_{1\leq i\leq n}\rho_{n_i}(x)}{(1+\varepsilon)+\rho_{n_i}(y)} \\
\text{(lower bound of } \rho_{n_i}(y)) &\geq \frac{1}{2^{n_i}} \cdot \frac{\varepsilon-\max\limits_{1\leq i\leq n}\rho_{n_i}(x)}{1+2\varepsilon} \\
\text{(removing dependence on } n_i) &\geq 2^{-\max\{n_1,\ldots,n_k\}} \cdot \frac{\varepsilon-\max\limits_{1\leq i\leq n}\rho_{n_i}(x)}{1+2\varepsilon}
\end{aligned}$$

It then follows that

$$d(x,z) < 2^{-\max\{n_1,\ldots,n_k\}} \cdot \frac{\varepsilon-\max\limits_{1\leq i\leq n}\rho_{n_i}(x)}{1+2\varepsilon}$$

implies $\rho_{n_i}(z) < \varepsilon$ for all $1 \leq i \leq n$, hence $z \in V(n_1,\ldots,n_k;\varepsilon)$. Therefore, $V(n_1,\ldots,n_k;\varepsilon)$ is open in X_d.

($\Leftarrow$) By the translation invariance of the two topologies at hand, it suffices to show that $B_\varepsilon(0)$ is open in X_ρ. For each $x \in B_\varepsilon(0)$, we set

$$r = \frac{1}{2}\min\{d(x,0), |\varepsilon - d(x,0)|\}$$

and note that $B_r(x) \subseteq B_\varepsilon(0)$. We find a natural number k such that $2^{-k} < r/2$ and observe that $\rho_i(x-y) < r/k$ implies

$$d(x,y) < \sum_{i=1}^{k} \frac{1}{2^n} \frac{\rho_i(x-y)}{1+\rho_i(x-y)} + \frac{1}{2^k} < \frac{1}{2}\sum_{i=1}^{k}\rho_i(x-y) + \frac{r}{2} < \frac{1}{2}\cdot k \cdot \frac{r}{k} + \frac{r}{2} = r,$$

whence $V(1,\ldots,k;r/k)$ is an X_ρ-open neighborhood of x in $B_r(x)$, hence in $B_\varepsilon(0)$. It follows that $B_\varepsilon(0)$ is open in X_ρ. □

Proposition 15.9. *A sequence $(x_n)_{n=1}^\infty$ in a topological vector space X generated by a countable family of seminorms $\{\rho_m\}_{m\in\mathbb{N}}$ converges to $x \in X$ if and only if $\rho_m(x_n - x) \to 0$ for each fixed $m \in \mathbb{N}$.*

Proof. This is a trivial consequence of Proposition 15.8. □

Proposition 15.10. *Let $\{\rho_m\}_{m\in\mathbb{N}}$ be a countable family of seminorms that generates the topology on a topological vector space X. If every sequence $(x_n)_{n=1}^\infty$ such that $\rho_m(x_{n_1} - x_{n_2}) \to 0$ as $n_1, n_2 \to \infty$ for each fixed $m \in \mathbb{N}$ converges to a limit in X, then X is a Fréchet space.*

Proof. By Proposition 15.9, the hypothesis translates to the metric

$$d(x,y) = \sum_{n=1}^\infty \frac{1}{2^n}\frac{\rho_n(x-y)}{1+\rho_n(x-y)}$$

being complete, whence the desired result follows from Proposition 15.8. □

We now consider two of the most important examples of Fréchet spaces.

Example 15.11. The space $\mathscr{C}^\infty(K)$ of smooth functions on a compact subset K of $\mathbb{R}$ comes with the natural seminorms

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

which are complete in the sense of Proposition 15.10. Similarly, the *Schwartz space* $\mathscr{S}(\mathbb{R})$ of smooth functions $f : \mathbb{R} \to \mathbb{C}$ such that the quantity

$$\rho_{n,m}(f) = \sup_{x\in\mathbb{R}} |x^m f^{(n)}(x)|$$

is finite for each $m, n \in \mathbb{N}$ is a Fréchet space with the seminorms $\rho_{n,m}$. These are examples of spaces of *test functions*, which forms the basis of distribution theory.

We conclude this section by placing some of the major theorems established in this chapter in their proper contexts. The omitted proofs can be found in many of the standard functional analysis textbooks.

Theorem 15.12 (Banach-Schauder, open mapping theorem)**.** *If $T : X \to Y$ is a surjective continuous linear operator between Fréchet spaces X and Y, then T is open.*

Theorem 15.13 (Closed graph theorem)**.** *A linear operator $T : X \to Y$ between Fréchet spaces X and Y is bounded if and only if $\Gamma(T)$ is closed.*

Theorem 15.14 (Banach-Steinhaus-Dieudonné, uniform boundedness principle)**.** *Let X be a Fréchet space, Y a normed linear space. Let Γ be a nonempty index set, and, for each $\alpha \in \Lambda$, we let $T_\alpha : X \to Y$ be a continuous linear operator. If*

$$\sup_{\alpha\in\Lambda} \|T_\alpha x\| < \infty$$

for all $x \in X$, then $\{T_\alpha\}_{\alpha\in\Lambda}$ is equicontinuous, viz., for each $\varepsilon > 0$, there exists a $\delta > 0$ such that $d(x,y) < \delta$ implies $\|T_\alpha x - T_\alpha y\| < \varepsilon$ for all $x, y \in X$ and $\alpha \in \Lambda$.

Theorem 15.15 (Analytic Hahn-Banach)**.** *Let X be a locally convex space, M a subspace of X, and l a continuous linear functional on M. Then there exists a continuous extension L of l onto the whole space X.*

Theorem 15.16 (Existence of nontrivial annihilators)**.** *If X is a locally convex space and M a closed proper subspace of X, then there exists a nonzero continuous linear functional L on X such that $L|_M = 0$.*

Theorem 15.17 (Hyperplane separation theorem)**.** *Let X be a real locally convex space and A and B nonempty, convex, disjoint subsets of X. If A is compact and B is closed, then there exists an $L \in X^*$ and an $\alpha \in \mathbb{R}$ such that*

$$\sup_{x \in A} L(x) < \alpha < \inf_{y \in B} L(y).$$

In other words, the hyperplane $\{z : L(z) = \alpha\}$ separates A and B.

Theorem 15.18 (Krein-Milman)**.** *If X is a locally convex space and K a nonempty, compact, convex subset of X, then $K = \overline{\text{co}(\text{ext}(A))}$.*

Theorem 15.19 (Banach-Alaoglu)**.** *Let X be a topological vector space. If K is the collection of continuous linear functionals on X that are bounded by 1 on a fixed neighborhood of 0 in X, then K is weak-$*$ compact.*

16. Hilbert Spaces

We now turn to the study of Hilbert spaces, a topological vector space that retains many geometric properties of the Euclidean space.

16.1. **Inner Products.** Recall that two vectors v and w in $\mathbb{R}^n$ are *orthogonal* if their dot product $v \cdot w = \sum v_i w_i$ is zero. If we are to speak of orthogonality in a broader context, it is first necessary to generalize the dot product.

Definition 16.1. An *inner-product space* is a complex vector space X with a map $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{C}$ such that

(i) $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$,
(ii) $\langle y, x \rangle = \overline{\langle x, y \rangle}$,
(iii) $\langle x, x \rangle \geq 0$, and
(iv) $\langle x, x \rangle = 0$ if and only if $x = 0$

for all $\alpha, \beta \in \mathbb{C}$ and $x, y, z \in V$.

(i) and (ii) imply that $\langle \cdot, \cdot \rangle$ is a *Hermitian form*, and (iii) implies that $\langle \cdot, \cdot \rangle$ is a *positive-definite form*; a positive-definite Hermitian form is a *semi-inner product*. If X is a real vector space, then Hermiticity reduces to *bilinearity*, i.e.,

$$\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle \text{ and } \langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle,$$

and *symmetry*

$$\langle x, y \rangle = \langle y, x \rangle.$$

Therefore, a semi-inner product on a real vector space is a positive-definite, symmetric bilinear form, and (iv) turns it into a *real inner product.* We prefer to work with complex inner products, as they possess certain advantages over real inner products that we shall see in later sections.

By the positive definiteness of the inner product, we see that

$$\|x\| = \sqrt{\langle x, x \rangle}$$

is nonnegative, and hermiticity implies that it is homogeneous just like a norm. We shall show that $\|\cdot\|$ is indeed a norm in due course, but, for now, we harness the basic properties of the inner product $\langle\cdot,\cdot\rangle$ and the associated "norm" $\|\cdot\|$. The first is a simple computational principle that directly generalizes the law of cosines in $\mathbb{R}^2$.

Proposition 16.2 (Generalized law of cosines)**.** *For all $x, y \in X$,*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x, y\rangle.$$

Proof. Hermiticity. □

In classical Euclidean geometry, the Pythagorean theorem could be deduced as a special case of the law of cosines in which the two sides are perpendicular to each other. We now generalize the notion of perpendicularity and use it to establish a Pythagorean identity for inner-product spaces.

Definition 16.3. Two vectors $x, y \in X$ are *orthogonal* if $\langle x, y\rangle = 0$. We write $x \perp y$ to denote that x and y are orthogonal.

Proposition 16.4 (Generalized Pythagorean theorem)**.** *For all $x_1, \ldots, x_n \in X$ such that $x_i \perp x_j$ for all $1 \leq i < j \leq n$,*

$$\left\|\sum_{i=1}^n x_i\right\|^2 = \sum_{i=1}^n \|x_i\|^2.$$

Proof. For $n = 2$, this is a special case of the generalized law of cosines. For $n > 2$, it suffices to apply the $n = 2$ case repeatedly. □

The next result is the foundational inequality of Schwarz, which generalizes Cauchy's inequality in $\mathbb{R}^n$.

Proposition 16.5 (Schwarz's inequality)**.** $|\langle x, y\rangle| \leq \|x\|\|y\|$ *for all $x, y \in X$, and the equality holds if and only if x is a scalar multiple of y.*

The key idea is to decompose x into two orthogonal vectors and to obtain an estimate via the Pythagorean theorem. This decomposition is (almost) always possible, and we take this up in Theorem 16.24.

Proof. We assume without loss of generality that $y \neq 0$ and set $\hat{y} = \|y\|^{-1}y$, so that $\|\hat{y}\| = 1$. We let

$$u = \langle x, \hat{y}\rangle\hat{y} \quad \text{and} \quad v = x - \langle x, \hat{y}\rangle\hat{y},$$

so that $x = u + v$ and $u \perp v$. By the Pythagorean theorem, we have

$$\|x\|^2 = \|u\|^2 + \|v\|^2 = \|\langle x, \hat{y}\rangle\hat{y}\|^2 + \|x - \langle x, \hat{y}\rangle\hat{y}\|^2.$$

Combined with the positive-definiteness of the inner product, the above identity implies that

$$\|x\|^2 \geq \|\langle x, \hat{y}\rangle\hat{y}\|^2 = |\langle x, \hat{y}\rangle|^2\|\hat{y}\|^2 = |\langle x, \hat{y}\rangle|^2,$$

where the equality holds if and only if $\|x - \langle x, \hat{y}\rangle\hat{y}\|^2$. The inequality is precisely Schwarz's inequality; the equality condition is equivalent to the attainment of the identity

$$x = \langle x, \hat{y}\rangle\hat{y} = \left(\frac{\langle x, \|y\|^{-1}y\rangle}{\|y\|}\right)y = \|y\|^{-2}\langle x, y\rangle y. \tag{16.6}$$

Therefore, if the equality in Schwarz's inequality holds, then (16.6) holds, whence, in particular, x is a scalar multiple of y. Conversely, if $x = \lambda y$, for some scalar λ, then

$$\|y\|^{-2}\langle x, y\rangle y = \|y\|^{-2}\langle \lambda y, y\rangle y = \lambda y = x,$$

which is precisely (16.6). □

With Schwarz's inequality at hand, we can now deduce that $\|\cdot\|$ is indeed a norm.

Proposition 16.7. *$\|x\| = \sqrt{\langle x, x\rangle}$ is a norm on X.*

Proof. Homogeneity follows hermiticity; positive-definiteness is inherited. To verify the triangle inequality, we observe that

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\,\mathrm{Re}\langle x, y\rangle$$

by the law of cosines. Applying Schwarz's inequality, we see that

$$2\,\mathrm{Re}\langle x, y\rangle = \langle x, y\rangle + \overline{\langle x, y\rangle} \le \|x\|\|y\| + \overline{\|x\|\|y\|} = 2\|x\|\|y\|,$$

whence we have

$$\|x + y\|^2 \le (\|x\| + \|y\|)^2,$$

as desired. □

We now define the single most important object in functional analysis.

Definition 16.8. A *Hilbert space* is a complete inner-product space, in the sense that the norm associated with the inner product induces a complete metric.

Example 16.9. The Euclidean spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual dot products are Hilbert spaces.

Example 16.10. The space $L^2(X, \mu)$ of square-integrable functions on a measure space (X, μ) with the inner product

$$\langle f, g\rangle_2 = \int_X f\overline{g}\,d\mu$$

is a Hilbert space.

Example 16.11. Let $\mathcal{T}$ be the set of all *trigonometric polynomials*

$$p(x) = \sum_{n=1}^{N} p_n e^{2\pi i n x}$$

on $[0, 1]$, with the inner product

$$\langle p, q\rangle = \sum_{n=1}^{N} p_n \overline{q_n}.$$

is an inner-product subspace of $L^2([0, 1])$. The space $\mathcal{T}_N$ of trigonometric polynomials of degree at most N is a closed subspace of $\mathcal{T}$ and hence a Hilbert subspace of $L^2([0, 1])$.

Example 16.12. Let $K \in \mathscr{C}_c(\mathbb{R}^2)$ be nonnegative and *symmetric*, i.e., $K(x,y) = K(y,x)$. The *weighted L^2 inner product*

$$\langle f, g\rangle_K = \iint f(x)\overline{g(y)}K(x,y)\,dx\,dy$$

is a complete inner product on $L^2(\mathbb{R}^2)$. Here we only show that

$$\|f\|_K = \left(\iint |f(x)|^2 K(x,y)\,dx\,dy\right)^{1/2}$$

is a complete norm. To this end, we let $(f_n)_{n=1}^\infty$ be a Cauchy sequence with respect to $\|\cdot\|_K$ and extract a subsequence $(f_{n_k})_{k=1}^\infty$ such that $\|f_{n_k} - f_{n_{k+1}}\|_K < 2^{-k}$. This subsequence converges pointwise almost everywhere to

$$f(x) = f_{n_1}(x) + \sum_{k=1}^\infty (f_{n_{k+1}}(x) - f_{n_k}(x)),$$

which is in $L^2(\mathbb{R}^2)$. Now, $|f_{n_k} - f|^2 K \in L^1(\mathbb{R}^2)$, and so

$$\begin{aligned}
\lim_{k\to\infty} \|f_{n_k} - f\|_K &= \lim_{k\to\infty} \iint |f_{n_k}(x,y) - f(x,y)|^2 K(x,y)\,dx\,dy \\
&= \iint \lim_{k\to\infty} |f_{n_k}(x,y) - f(x,y)|^2 K(x,y)\,dx\,dy \\
&= 0
\end{aligned}$$

by the dominated convergence theorem. The desired result now follows.

16.2. **Orthogonality.** Having reviewed the definition and basic examples of a Hilbert space, we now turn to the promised discussion on orthogonality. The next result encodes the orthogonality property of inner product spaces in a norm identity, attributed to Pascual Jordan and John von Neumann in [Tes09].

Theorem 16.13 (Parallelogram law / Polarization identity)**.** *If X is an inner-product space, then the* **parallelogram law**

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

and the **polarization identity**

$$\langle x, y\rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2\right)$$

hold. Conversely, if X is a normed linear space that satisfies the parallelogram law, then the quadratic form given by the polarization identity is an inner product on X.

Proof. If X is an inner-product space, then the parallelogram law follows from the law of cosines (Proposition 16.2) and the polarization identity from the hermiticity of the inner product. Conversely, we assume that X is a normed linear space satisfying the parallelogram law and consider the quadratic form $\langle\cdot,\cdot\rangle$ given by the polarization identity. Observe that

$$\begin{aligned}
4\langle x, y\rangle &= \|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2 \\
&= \|y+x\|^2 - \|y-x\|^2 + i\|y-ix\|^2 - i\|y+ix\|^2 \\
&= \overline{(\|y+x\|^2 - \|y-x\|^2 - i\|y-ix\|^2 + i\|y+ix\|^2)} \\
&= 4\overline{\langle y, x\rangle}.
\end{aligned}$$

In particular, $\langle x, x\rangle = \overline{\langle x, x\rangle}$, and so

$$4\langle x, x\rangle = 4\,\mathrm{Re}\langle x, x\rangle = \|x+x\|^2 + \|x-x\|^2 = 4\|x\|^2,$$

whence $\langle\cdot,\cdot\rangle$ is positive-definite. In particular, $\langle x, x\rangle = 0$ if and only if $x = 0$.

It remains to show that $\langle\cdot,\cdot\rangle$ is linear in the first variable. By the parallelogram law, we have

$$\begin{aligned} & 4(\langle x, z\rangle + \langle y, z\rangle) \\ = \; & \|(x+y)+2z\|^2 - \|(x+y)-2z\|^2 + i\|(x+y)+2iz\|^2 - \|(x+y)-2iz\|^2 \\ = \; & 8(\langle (x+y)/2, z\rangle). \end{aligned}$$

In particular, $\langle x, z\rangle = \frac{1}{2}\langle 2x, z\rangle$, and so

$$\langle x, z\rangle + \langle y, z\rangle = 2\langle (x+y)/2, z\rangle = \langle x+y, z\rangle. \tag{16.14}$$

Iterating these two identities imply that

$$\sum_{n=-N}^{N} a_n 2^n \langle x, z\rangle = \left\langle \left(\sum_{n=-N}^{N} a_n 2^n\right) x, z \right\rangle \tag{16.15}$$

for all $N \in \mathbb{N}$ and $a_n \in \mathbb{N}$.

Now, the norm is continuous, and so $\langle\cdot,\cdot\rangle$ is continuous in the first variable. If this claim holds, then (16.15) implies that

$$r\langle x, z\rangle = \langle rx, z\rangle \tag{16.16}$$

for all $r \geq 0$. Simple computations show that $-\langle x, z\rangle = \langle -x, z\rangle$ and $i\langle x, z\rangle = \langle ix, z\rangle$, whence (16.16) now implies that

$$\lambda\langle x, z\rangle = \langle \lambda x, z\rangle.$$

This, combined with (16.14), establishes the hermiticity of $\langle\cdot,\cdot\rangle$, thereby showing that $\langle\cdot,\cdot\rangle$ is an inner product. □

The continuity property used in the above proof holds in general:

Proposition 16.17. *The inner product is continuous as a map from $\langle\cdot,\cdot\rangle$ into $\mathbb{C}$.*

Proof. If $x_n \to x$ and $y_n \to y$, then

$$\lim_{n\to\infty} |\langle x, y\rangle - \langle x_n, y_n\rangle| \leq \lim_{n\to\infty} \|x - x_n\|\|y - y_n\| = 0$$

by Schwarz's inequality (Proposition 16.5). □

With the parallelogram law and the polarization identity at hand, we can now construct the *Hilbert completion* of an inner-product space. Given an inner-product space X, we take its Banach completion (Theorem 8.13) $\tilde{X}$. Since the norm is continuous, the parallelogram law on X evidently carries over to $\tilde{X}$, whence the inner product given by the polarization identity turns $\tilde{X}$ into a Hilbert space. In light of this, analysts often refer to inner-product spaces as *pre-Hilbert spaces*.

Example 16.18. $\mathscr{C}^1[0,1]$ with the inner product

$$\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx + \int_0^1 f'(x)\overline{g'(x)}\,dx$$

is a pre-Hilbert space. Its Hilbert completion is the *Sobolev space of order 1*, denoted by H^1.

Example 16.19. While all norms on $\mathbb{R}^n$ are equivalent (Theorem 8.18), not too many of them admit an inner product. For example, the maximum norm

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

does not satisfy the parallelogram law. Indeed,

$$\|(2,0)+(0,1)\|_\infty^2 + \|(2,0)-(0,1)\|_\infty^2 = 8 \neq 10 = 2(\|(2,0)\|_\infty^2 + \|(0,1)\|_\infty^2),$$

and this counterexample can easily be extended to handle the $\mathbb{R}^n$ case.

We now exploit the orthogonality property to study the geometry of Hilbert spaces. First, we generalize the notion of *vectors perpendicular to a plane* used in the study of three-dimensional Euclidean geometry.

Definition 16.20. Given a subset E of an inner product space V, we define the *orthogonal complement of* E to be the set

$$E^\perp = \{x \in V : x \perp y \text{ for all } y \in E\}.$$

Recall that a *projection* on a vector space V is an idempotent linear endomorphism, i.e., a linear transformation $P : V \to V$ such that $P^2 = P$. Note that $P|_{\operatorname{im} P}$ is the identity operator I, and that we have the direct-sum decomposition

$$V = P(V) \oplus (I-P)(V),$$

viz., each $x \in V$ can be written as the sum $x = y + z$ with $y \in P(V)$ and $z \in (I-P)(V)$ in precisely one way: namely, $y = Px$ and $z = (I-P)x = x - Px$. In view of the direct sum decomposition, we often refer to P as *the projection of* V *onto* $P(V)$ *along* $(I-P)(V)$.

One of the most important results in Hilbert-space theory is the existence of *orthogonal projections*:

Definition 16.21. Let M be a closed linear subspace of $\mathcal{H}$. The *orthogonal projection* of $\mathcal{H}$ onto M is the projection P of $\mathcal{H}$ onto M along $M^\perp$.

Remark 16.22. P is an orthogonal projection if and only if $(\operatorname{im} P)^\perp = \ker P$.

Lemma 16.23. *If E is a subset of an inner-product space X, then $E^\perp$ is a closed linear subspace of X.*

Proof of lemma. It is clear that $E^\perp$ is a linear subspace of X. To show that $E^\perp$ is closed, it suffices to note that

$$E^\perp = \bigcap_{y\in E} \{y\}^\perp,$$

and that each $\{y\}^\perp$ is the preimage of $\{0\}$ of the continuous mapping $x \mapsto \langle x, y\rangle$ (Proposition 16.17). □

Theorem 16.24 (Existence of orthogonal projections)**.** *Let $\mathcal{H}$ be a Hilbert space. If M is a closed, convex subset of $\mathcal{H}$, then there exists a linear operator each $x \in \mathcal{H}$ admits precisely one $x_M \in M$ such that*

$$\|x - x_M\| = \inf_{y\in M} \|x - y\|.$$

If M is, in addition, a linear subspace of $\mathcal{H}$, then there exists a projection P of $\mathcal{H}$ along $M^\perp$ onto M, called the orthogonal projection *of $\mathcal{H}$ onto M. In particular,*

$$\mathcal{H} = M \oplus M^\perp.$$

We remark that the hypothesis that M is a closed *convex* linear subspace is superfluous, as every linear subspace is convex.

Proof of theorem. Assume for now that M is a closed, convex subset of $\mathcal{H}$. Let $(y_n)_{n=1}^{\infty}$ be a sequence in M such that

$$\lim_{n\to\infty} \|x - y_n\| = d(x, M) = \inf_{y\in M} \|x - y\|.$$

Since M is convex, $(y_n + y_m)/2 \in M$, and so $\|x - (y_n + y_m)/2\| \geq d(x, M)$. The parallelogram law (Theorem 16.13) implies that

$$\begin{aligned} 4d(x,M)^2 &\leq 4\|x - (y_n + y_m)/2\|^2 \\ &\leq 4\|x - (y_n + y_m)/2\|^2 + \|y_n - y_m\|^2 \\ &= \|(x - y_m) + (x - y_n)\|^2 + \|(x - y_m) - (x - y_n)\|^2 \\ &= 2(\|x - y_m\|^2 + \|x - y_n\|^2), \end{aligned}$$

whence

$$\begin{aligned} \lim_{n,m\to\infty} \|y_n - y_m\|^2 &\leq \lim_{n,m\to\infty} 2(\|x - y_m\|^2 + \|x - y_n\|^2) - 4d(x,M)^2 \\ &= 2(d(x,M)^2 + d(x,M)^2)) - 4d(x,M)^2 \\ &= 0. \end{aligned}$$

It follows that $(y_n)_{n=1}^{\infty}$ is Cauchy.

We invoke the completeness of $\mathcal{H}$ to find $x_M = \lim_n y_n$. By construction, $\|x - x_M\| = d(x, M)$. To show that the minimizer is unique, we let $y \in M$ be another minimizer, i.e., $\|x - y\| = d(x, M)$. This, in particular, implies that $\|x - (x_M + y)/2\| = d(x, M)$, and so

$$\begin{aligned} 4d(x,M)^2 &= 2(d(x,M)^2 + d(x,M)^2) \\ &= 2(\|x - x_M\|^2 + \|x - y\|^2) \\ &= \|(x - x_M) + (x - y)\|^2 + \|(x - x_M) - (x - y)\|^2 \\ &= 4\|x - (x_M + y)/2\|^2 + \|y - x_M\|^2 \\ &= 4d(x,M)^2 + \|y - x_M\|^2 \end{aligned}$$

by the parallelogram law (Theorem 16.13). It follows that $\|y - x_M\|^2 = 0$, whence $y = x_M$.

We now assume that M is a closed linear subspace. Define an operator $P : \mathcal{H} \to \mathcal{H}$ by setting $Px = x_M$ for each $x \in \mathcal{H}$. We claim that P is a projection of $\mathcal{H}$ along $M^\perp$ onto M. To this end, we first note that $P^2x = (x_M)_M$ is evidently x_M, for the unique distance minimizer for any element of M is the element itself. Therefore, $P^2 = P$, and the argument also shows that $M \subseteq P(M)$. and P is a projection. Secondly, we have shown above that the distance minimizer is always in M, and so $P(M) \subseteq M$. It follows that P is a projection onto M.

Lastly, we must check that $(I - P)x = x - x_M$ is an element of $M^\perp$, which is a closed linear subspace by Lemma 16.23. This amounts to showing that $\langle x - x_M, y\rangle$ vanishes for all $y \in M^\perp$. To this end, we fix $y \in M^\perp$ and consider $f : \mathbb{R} \to [0, \infty)$ given by the formula

$$f(t) = \|x - x_M + ty\|^2.$$

By the law of cosines (Proposition 16.2), we can write

$$f(t) = \|x - x_M\|^2 + t^2\|y\|^2 + 2t \operatorname{Re}\langle x - x_M, y\rangle.$$

Now, we observe that

$$f(t) = \|x - (x_M - ty)\|^2$$

and $x_M - ty \in M$, whence the distance function f is minimized when $x_M - ty = x_M$, or $t = 0$. Therefore,

$$\|x - x_M\|^2 = f(0) \leq f(t) = \|x - x_M\|^2 + t^2\|y\|^2 + 2t\,\mathrm{Re}\langle x - x_M, y\rangle$$

for all $t \in \mathbb{R}$. This is only possible if $\langle x - x_M, y\rangle = 0$. It now follows that P is a projection along $M^\perp$, thus completing the proof. □

An immediate corollary of the direct-sum decomposition is that proper closed subspaces must have a nontrivial orthogonal complement.

Corollary 16.25 (Existence of nontrivial orthogonal complements)**.** *If M is a closed proper linear subspace of $\mathcal{H}$, then there exists a nonzero vector $y \in \mathcal{H}$ such that $\langle x, y\rangle = 0$ for all $x \in M$.* □

For each $u \in M^\perp$, we obtain a linear functional $l_u : \mathcal{H} \to \mathbb{C}$ given by $l_u(x) = \langle x, u\rangle$. By the continuity of the inner product (Proposition 16.17), l_u is bounded. Furthermore, l_y vanishes on M, whence l_u is an an element of the *annihilator* of M, viz., the collection of bounded linear functionals on that vanish on M. In this sense, Corollary 16.25 can be thought of as the existence of nontrivial annihilators that we have established in Corollary 11.13. Direct comparisons bewteen orthogonal complements and annihilators are made in Appendix A, Section 19.2.

The construction of the bounded linear functional l_u evidently applies to any $u \in \mathcal{H}$, and this shows that there are plenty of bounded linear functionals on $\mathcal{H}$. Shockingly, they fill up $\mathcal{H}^*$, as we shall see in the following result.

Theorem 16.26 (Riesz representation)**.** *Every bounded linear functional $l : \mathcal{H} \to \mathbb{C}$ can be written as an action of the inner product against a vector $u \in \mathcal{H}$, i.e.,*

$$lx = \langle x, u\rangle$$

for all $x \in \mathcal{H}$. Furthermore, this u is unique, and $\|l\| = \|u\|$, whence $\mathcal{H}$ is isometrically isomorphic to $\mathcal{H}^*$ and is thus reflexive.

We remark that the natural notion of isomorphisms between Hilbert spaces is that of a *unitary isomorphism*:

Definition 16.27. A linear operator $T : X \to Y$ between two inner-product spaces is *unitary* if $\langle Tx, Ty\rangle = \langle x, y\rangle$ for all $x, y \in X$. T is a *unitary isomorphism* if T is a unitary operator and a linear isomorphism.

Nevertheless, the polarization identity (Theorem 16.13) implies that a linear isomorphism between pre-Hilbert spaces is unitary if and only if it is isometric. In practice, therefore, we never have to check whether the inner product is preserved.

Proof of theorem. If l is the zero functional, then 0 is clearly the only vector that does the job. If l is nontrivial, then $\ker l$ is a closed proper linear subspace of $\mathcal{H}$, whence $(\ker l)^\perp$ is nontrivial by Corollary 16.25. We pick a unit vector $z \in (\ker l)^\perp$ and set $u = (\overline{lz})z$. Note that $(lx)z - (lz)x \in \ker l$ for all $x \in \mathcal{H}$, and so

$$0 = \langle (lx)z - (lz)x, z\rangle = (lx)\langle z, z\rangle - \langle x, (\overline{lz})z\rangle = lx - \langle x, u\rangle,$$

or $lx = \langle x, u\rangle$. Finally, we observe that

$$\|u\| = |\overline{(lz)}| = |lz| \leq \|l\|\|z\| = \|l\|$$

and

$$|lz| = |\overline{lz}| \|z\| = \|u\| \|z\|;$$

the latter implies that $\|u\| \geq \|l\|$, whence $\|u\| = \|l\|$. □

We pause to reflect on a particularly useful consequence of the orthogonality property of the Euclidean space: $\mathbb{R}^n$ is equipped with n mutually orthogonal axes that admit a particularly convenient coordinate representation of n-vectors. Our next goal is to recover this coordinate representation in the Hilbert-space setting. The proper way to set up "mutually orthogonal axes" is via an orthonormal set vectors.

Definition 16.28. A set of vectors $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ in an inner-product space X is *orthonormal* if $\|u_\alpha\| = 1$ for all $\alpha \in \mathcal{I}$ and $u_\alpha \perp u_\beta$ for all $\alpha \neq \beta$.

The next result shows that a finite orthonormal set is, in general, not sufficient for a full-blwon coordinate representation.

Proposition 16.29 (Finite Bessel's inequality)**.** *If $\{u_1, \ldots, u_N\}$ is an orthonormal set in an inner-product space X, then*

$$\sum_{n=1}^{N} |\langle x, u_n \rangle|^2 \leq \|x\|^2$$

for all $x \in X$, and the equality holds if and only if

$$x = \sum_{n=1}^{N} \langle x, u_n \rangle u_n.$$

Proof. Since $(\sum \langle x, u_n \rangle u_n)$ is orthogonal to $x - \sum \langle x, u_n \rangle$, the Pythagorean theorem (Proposition 16.4) implies that

$$\|x\|^2 = \left\| \sum_{n=1}^{N} \langle x, u_n \rangle u_n \right\|^2 + \left\| x - \sum_{n=1}^{N} \langle x, u_n \rangle u_n \right\|^2.$$

Now, the vectors $\langle x, u_1 \rangle u_1, \ldots, \langle x, u_n \rangle u_n$ are mutually orthogonal, and so another application of the Pythagorean theorem yields

$$\begin{aligned} \|x\|^2 &= \sum_{n=1}^{N} \|\langle x, u_n \rangle u_n\|^2 + \left\| x - \sum_{n=1}^{N} \langle x, u_n \rangle u_n \right\|^2 \\ &= \sum_{n=1}^{N} |\langle x, u_n \rangle|^2 + \left\| x - \sum_{n=1}^{N} \langle x, u_n \rangle u_n \right\|^2. \end{aligned}$$

The desired result now follows. □

We now recall that the sum of an infinite set $\{r_\alpha\}_{\alpha \in \mathcal{I}}$ of nonnegative real numbers is defined as the supremum of all the finite partial sums. If $\sum_\alpha r_\alpha$ is finite, then all but countably many elements in $\{r_\alpha\}$ must be zero, for otherwise we can extract a monotonically increasing sequence $(r_n)_{n=1}^{\infty}$ in $\{r_\alpha\}$. With this observation, we can extend finite Bessel's inequality to infinite orthonormal sets.

Corollary 16.30 (Bessel's inequality)**.** *If $\{u_\alpha\}_{\alpha \in \mathcal{I}}$ is an orthonormal set in an inner-product space, then*

$$\sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2 \leq \|x\|^2$$

for all $x \in X$. In particular, $\{\alpha : \langle x, u_\alpha \rangle \neq 0\}$ is a countable set. □

In order to recover the equality condition we have established for finite Bessel's inequality, we must deal with infinite sums of vectors. The appropriate notion of convergence for us is that of *unconditional convergence*: a countable sum

$$\sum_{n=1}^{\infty} x_n$$

of vectors in a topological vector space *converges uncounditionally* if there exists a vector x such that the sequence of partial sums converges to x for all possible reordering of indices. Note that unconditional convergence of vectors in a normed linear space is *not* equivalent to the absolute convergence

$$\sum_{n=1}^{\infty} \|x_n\| < \infty;$$

in fact, a theorem of Dvoretzky and Rogers ([DR50], Theorem 1) shows that unconditional convergence implies absolute convergence in a Banach space if and only if the space is finite-dimensional.

With this in mind, we deal with the task of recovering the equality condition:

Proposition 16.31 (Definition of orthonormal basis)**.** *Let $\mathscr{U} = \{u_\alpha\}_{\alpha \in \mathcal{I}}$ be an orthonormal set in a Hilbert space $\mathcal{H}$. The following are equivalent:*

(a) $\mathscr{U}$ is an **orthonormal basis***, viz.,*

$$x = \sum_{\alpha \in \mathcal{I}} \langle x, u_\alpha \rangle u_\alpha$$

for all $x \in \mathcal{H}$, where the sum converges unconditionally;

(b) **Parseval's identity** *holds for $\mathscr{U}$, viz.,*

$$\|x\|^2 = \sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2$$

for all $x \in \mathcal{H}$.

(c) $\mathscr{U}$ is **complete***, i.e. $\mathscr{U}^\perp = \{0\}$.*

Proof. $(a) \Rightarrow (b)$. By Bessel's inequality, the set $\mathcal{I}_x = \{\alpha \in \mathcal{I} : \langle x, u_\alpha \rangle \neq 0\}$ is countable. We index the elements of $\mathcal{I}_x$ by natural numbers: $\mathcal{I}_x = \{\alpha_n\}_{n \in \mathbb{N}}$. Parseval's identity is now obtained by considering the limiting case of the identity

$$\|x\|^2 = \sum_{n=1}^{N} |\langle x, u_n \rangle|^2 + \left\| x - \sum_{n=1}^{N} \langle x, u_n \rangle u_n \right\|^2$$

from the proof of finite Bessel's inequality.

$(b) \Rightarrow (c)$. If $\mathscr{U}^\perp$ contains a nonzero vector x, then

$$\|x\|^2 > 0 = \sum_{\alpha \in \mathcal{I}} |\langle x, u_\alpha \rangle|^2.$$

$(c) \Rightarrow (a)$. By Bessel's inequality, the set $\mathcal{I}_x = \{\alpha \in \mathcal{I} : \langle x, u_\alpha \rangle \neq 0\}$ is countable. We index the elements of $\mathcal{I}_x$ by natural numbers: $\mathcal{I}_x = \{\alpha_n\}_{n \in \mathbb{N}}$. Bessel's inequality implies that $\sum |\langle x, u_\alpha \rangle|^2 < \infty$, whence it follows from the Pythagorean theorem (Proposition 16.4) that

$$\lim_{N \to \infty} \left\| \sum_{n=N}^{N+m} \langle x, u_n \rangle u_n \right\|^2 = \lim_{N \to \infty} \sum_{n=N}^{N+m} |\langle x, u_n \rangle|^2 = 0$$

for all $m \in \mathbb{N}$. By completeness of $\mathcal{H}$, the series $\sum \langle x, u_n \rangle u_n$ converges. Now,

$$\left\langle x - \left(\sum_{n=1}^{\infty} \langle x, u_n \rangle u_n \right), u_n \right\rangle = \langle x, u_n \rangle - \langle x, u_n \rangle = 0$$

for all $n \in \mathbb{N}$, whence by the completness of $\mathscr{U}$ we conclude that $x - \sum \langle x, u_n \rangle u_n = 0$, or $x = \sum \langle x, u_n \rangle u_n$. □

With the new tools available, we can now establish swiftly the coordinate representation we have advertised.

Theorem 16.32. *Every Hilbert space has an orthonormal basis.*

Sketch of proof. Apply Zorn's lemma on the collection of orthonormal sets ordered by inclusion: the maximal element is complete. □

The existence of an orthonormal basis, combined with Parseval's identity, implies the following characterization of Hilbert spaces.

Corollary 16.33. *Every Hilbert space is unitarily isomorphic to $l^2(\mathcal{I})$ for some set $\mathcal{I}$ with the counting measure.*

Sketch of proof. Let $\mathcal{H}$ be a Hilbert space. By Theorem 16.32, $\mathcal{H}$ has an orthonormal basis $\{u_\alpha\}_{\alpha \in \mathcal{I}}$. For each $x \in \mathcal{H}$, we define $f_x \in l^2(\mathcal{I})$ by setting

$$f_x(\alpha) = \langle x, u_\alpha \rangle.$$

The mapping $x \mapsto f_x$ is then an isometric isomorphism. □

16.3. **Separability.** While theoretically sound, the coordinate presentation give by Theorem 16.32 is unwieldy in many cases. In practice, it is useful to have a countable orthonormal basis, so as to dispense with the infinite-sums-as-suprema business. To this end, we recall that a topological space is *separable* if it contains a countable dense subset. An orthonormal set $\mathscr{U}$ is an orthonormal basis if and only if the rational linear combination of $\mathscr{U}$ is dense, and so we expect to see a strong connection between separability and the existence of a countable orthonormal basis.

Theorem 16.34. *A Hilbert space $\mathcal{H}$ has a countable orthonormal basis if and only if $\mathcal{H}$ is separable.*

Proof. The equivalence is trivial if $\mathcal{H}$ is finite-dimensional, hence we assume that $\mathcal{H}$ is infinite-dimensional.

($\Leftarrow$) Let $\mathscr{U}$ be a countable orthonormal basis of $\mathcal{H}$ and take the collection $\mathscr{B}$ of rational linear combinations of $\mathscr{U}$. $\mathscr{B}$ is dense in $\operatorname{span} \mathscr{U}$, and $\mathscr{U}$ is dense in $\mathcal{H}$ by the completeness characterizaiton of orthonormal bases (Proposition 16.31). Since $\mathscr{B}$ is countable, it follows that $\mathcal{H}$ is separable.

($\Rightarrow$) Let $\{x_m\}_{m \in \mathbb{N}}$ be a dense subset of $\mathcal{H}$. We build a countable linearly independent subset $\{e_n\}_{n \in \mathbb{N}}$ of $\{x_m\}_{m \in \mathbb{N}}$ as follows: let $e_1 = x_1$ and $m_1 = 1$; for each $n > 1$, we let m_n be the smallest natural number bigger than m_{n-1} such that $\{e_1, \ldots, e_{n-1}, x_{m_{n-1}+1}, \ldots, x_{m_n}\}$ is linearly independent. This process must continue indefinitely, as $\operatorname{span}\{x_m\} = \operatorname{span}\{\, e_n\}$. In particular, $\overline{\operatorname{span}\{\, e_n\}} = \mathcal{H}$.

The bulk of the remaining work is to establish the infinite-dimensional generalization of a familiar tool from linear algebra:

Lemma 16.35 (Gram-Schmidt process)**.** *If $\{e_n\}_{n\in\mathbb{N}}$ is a linearly independent subset of $\mathcal{H}$, then there exists an orthonormal set $\{u_N\}_{N\in\mathbb{N}}$ such that $\operatorname{span}\{e_n\} = \operatorname{span}\{u_N\}$ and that each u_N is a linear combination of $e_1, \ldots, e_N$.*

Proof of lemma. We first note that $\|e_1\| > 0$ by the linear independence of $\{e_n\}$. Set

$$u_1 = \frac{e_1}{\|e_1\|}$$

and, for each $N > 1$, define

$$u_{N+1} = \frac{e_{N+1} - \sum_{n=1}^{N} \langle e_{N+1}, e_n\rangle e_n}{\left\| e_{N+1} - \sum_{n=1}^{N} \langle e_{N+1}, e_n\rangle e_n \right\|}.$$

At each step, the linear independence of $\{e_n\}$ implies that u_{N+1} is well-defined and $\{u_1, \ldots, u_{N+1}\}$ is an orthonormal set whose linear span agrees with the linear span of $\{e_1, \ldots, e_{N+1}\}$. The lemma now follows. □

We now observe that Proposition 16.23 implies

$$\{u_N\}^\perp = (\operatorname{span}\{u_N\})^\perp = (\operatorname{span}\{e_n\})^\perp = (\overline{\operatorname{span}\{e_n\}})^\perp = \mathcal{H}^\perp = \{0\},$$

whence by the completeness characterization of orthonormal bases (Proposition 16.31) $\{u_N\}$ is an orthonormal basis. □

Corollary 16.36. *All infinite-dimensional separable Hilbert spaces are unitarily isomorphic to one another.*

Proof. This follows at once from Theorem 16.34 and Corollary 16.33. □

Note that we do not need Zorn's lemma to develop the theory of separable Hilbert spaces, which, in turn, makes the theory more *constructive*. The reliance of the above collary on Theorem 16.32 (from which Corollary 16.33 follows) is merely for convenience. Duplicating the proof of Corollary 16.33 with Theorem 16.34 instead of 16.32 allows us to avoid Zorn's lemma.

We conclude the subsection by recording some examples of separable Hilbert spaces and their orthonormal bases.

Example 16.37 (Kronecker delta basis of l^2). Let

$$e_n(m) = \begin{cases} 1 & \text{if } n = m; \\ 0 & \text{otherwise} \end{cases}$$

for all $n, m \in \mathbb{Z}$. Then $\{e_n\}_{n\in\mathbb{Z}}$ is an orthonormal basis in l^2.

Example 16.38 (L^2 theory of Fourier series). The *Fourier basis* $\{e_n = e^{inx}\}_{n\in\mathbb{Z}}$ on $[-\pi, \pi]$ is an orthonormal basis on $L^2([-\pi, \pi])$. The basis property yields the *Fourier inversion formula*

$$f(x) = \sum_{n\in\mathbb{Z}} \langle f, e_n\rangle e_n = \sum_{n\in\mathbb{Z}} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-int}\, dt \right) e^{inx},$$

and Parseval's identity yields the *Plancherel identity*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x)|^2\,dx = \sum_{n\in\mathbb{Z}}\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt\right|^2.$$

Example 16.39 (One-dimensional dyadic harmonic analysis). A *dyadic interval* in $\mathbb{R}$ is an interval of the form

$$[m2^{-k},(m+1)2^{-k})$$

for some $m,k\in\mathbb{Z}$. Given a dyadic interval $I=[m2^{-k},(m+1)2^{-k})$, we let

$$I_L=[m2^{-k},(m+1/2)2^{-k}) \quad\text{and}\quad I_R=[(m+1/2)2^{-k},(m+1)2^{-k})$$

and define the *Haar function associated with the interval I*

$$h_I = m(I)^{-1/2}\left(\chi_{I_L}-\chi_{I_R}\right).$$

The collection

$$\mathscr{H}=\{h_I : I \text{ is a dyadic interval in } \mathbb{R}\}$$

is a countable orthonormal basis in $L(\mathbb{R})$, called the *Haar wavelet basis* of $L^2(\mathbb{R})$.

In general, $\varphi\in L^2(\mathbb{R})$ is called a *wavelet* if the set $\{\varphi_{m,k} : m,k\in\mathbb{Z}\}$ consisting of the functions

$$\varphi_{m,k}(x)=2^{-m/2}\varphi(2^m x-n)$$

is an orthonormal basis of $L^2(\mathbb{R})$. Observe that the Haar function $h_{[0,1)}$ is a wavelet. It can be shown that a smooth wavelet exists.

Dyadic harmonic analysis is deeply intertwined with Littlewood-Paley theory of frequency localization: see Chapter 8 of [MS13] or Chapter 4 of [Gra10a] for details.

16.4. **Algebra of Operators.** We conclude the section with a quick development of the algebra of bounded operators on a Hilbert space, which is a generalization of the algebra of matrices (Section 4).

For the remainder of this section, we fix a Hilbert space $\mathcal{H}$ over $\mathbb{C}$.

Definition 16.40. The *product* TU of two operators $T,U\in\mathscr{B}(\mathcal{H})$ is the composite function $T\circ U$.

We observe at once that $TU\in\mathscr{B}(\mathcal{H})$. The following basic properties of the multiplication operators are trivially verified.

Proposition 16.41. *Let $T,U,V\in\mathscr{B}(\mathcal{H})$ and $\lambda\in\mathbb{C}$. The multiplication operation satisfies the following properties:*
(a) **Associativity**. $(TU)V=T(UV)$;
(b) **Distributivity**. $T(U+V)=TU+TV$ *and* $(T+U)V=TV+UV$;
(c) **Bilinearity**. $\lambda(TU)=(\lambda T)U=T(\lambda U)$;
(d) **Existence of Unit**. $IU=UI=I$*, where I is the identity operator.*

The above properties turns $\mathscr{B}(\mathcal{H})$ into a *unital associative* $\mathbb{C}$*-algebra*:

Definition 16.42. A vector space $\mathscr{A}$ over $\mathbb{F}$[7] is an *associative algebra over* $\mathbb{F}$, or an *associative* $\mathbb{F}$*-algebra*, if there exists a multiplication map $\mathfrak{m}:\mathscr{A}\times\mathscr{A}\to\mathscr{A}$ that is

- *associative*, i.e., $x(yz)=(xy)z$,

[7]Of course, associative algebras over a ring R can be defined by considering R-moudles with an appropriate multiplication operation, but this level of generality is unnecessary for this course.

- *distributive*, i.e., $x(y+z) = xy + xz$ and $(x+y)z = xz + yz$, and
- *bilinear*, i.e., $\lambda(xy) = (\lambda x)y = x(\lambda y)$

for all $x, y, z \in \mathscr{A}$ and $\lambda \in \mathbb{F}$, where we have written xy to denote $\mathfrak{m}(x, y)$. An associative $\mathbb{F}$-algebra $\mathscr{A}$ is *unital* if there exists an element $1 \in \mathscr{A}$ such that

$$1x = x1 = 1$$

for all $x \in \mathscr{A}$, and *commutative* if

$$xy = yx$$

for all $x, y \in \mathscr{A}$.

Recall that every normed linear space is a topological vector space. Indeed, the vector addition operation and the scalar multiplication operation are continuous. It is then natural to expect that newly-defined multiplication operation is continuous as well—and it indeed is.

Proposition 16.43. $\|TU\| \leq \|T\|\|U\|$ *for all* $T, U \in \mathscr{B}(\mathcal{H})$.

In fact, we have already used this fact once in the proof of Proposition 11.18. This turns $\mathscr{B}(\mathcal{H})$ into a *unital Banach algebra*:

Definition 16.44. An associative $\mathbb{F}$-algebra $\mathscr{A}$ is a *Banach algebra* if there exists a norm $\|\cdot\|$ on $\mathscr{A}$ that satisfies the inequality

$$\|xy\| \leq \|x\|\|y\|$$

for all $x, y \in \mathscr{A}$. A Banach algebra $\mathscr{A}$ is *unital* or *commutative* if $\mathscr{A}$ as an associative algebra is unital or commutative, respectively.

We now generalize the conjugate transpose operation on $\mathscr{M}_n(\mathbb{F})$. Recall the definition of the *adjoint* $T^* : Y^* \to X^*$ of a bounded linear operator $T : X \to Y$ between normed linear space from Definition 11.17. If $X = Y = \mathcal{H}$, then X^* and Y^* are isometrically isomorphic to $\mathcal{H}$ by the Riesz representation theorem (Theorem 16.26), and so we could consider T^* as an element of $\mathscr{B}(\mathcal{H})$. Indeed, for each $l \in \mathcal{H}^*$, the Riesz representation theorem yields a unique $y \in \mathcal{H}$ such that $lx = \langle x, y \rangle$, and so

$$(T^*l)(x) = (lT)(x) = l(Tx) = \langle Tx, y \rangle.$$

By the Riesz representation theorem, there exists a unique $y^* \in \mathcal{H}$ such that $T^*l(x) = \langle x, y^* \rangle$, and so

$$\langle x, y^* \rangle = (T^*l)(x) = \langle Tx, y \rangle.$$

Considering T^* as the operator that sends y to y^*, we are led to the following definition:

Definition 16.45. The *Hilbert adjoint*, or *Hermitian adjoint*, of $T \in \mathscr{B}(\mathcal{H})$ is the operator $T^* : \mathcal{H} \to \mathcal{H}$ that satisfies the identity

$$\langle Tx, y \rangle = \langle x, T^*y \rangle$$

for all $x, y \in \mathcal{H}$.

By the above discussion, we see that the Hilbert adjoint of an operator is always well-defined. Since the conjugate transpose A^* of an n-by-n matrix A with complex entries satisfies the dot-product identity

$$(Ax) \cdot y = x \cdot (A^*y),$$

the Hilbert adjoint can be considered as a direct generalization of the conjugate transpose operation.

The following basic properties of the adjoint operation are easily verified.

Proposition 16.46. *Let $T, U \in \mathscr{B}(\mathcal{H})$ and $\lambda \in \mathbb{C}$.*

(a) $*$ is an **involution**: $(T^*)^* = T$.
(b) $*$ is **conjugate-linear**: $(T+U)^* = T^* + U^*$ *and* $(\lambda T)^* = \bar{\lambda} T^*$.
(c) $*$ is **contravariant**: $(TU)^* = U^* T^*$.

The above properties turn $\mathscr{B}(\mathcal{H})$ into a *unital Banach $*$-algebra*:

Definition 16.47. An *involution* on a Banach algebra $\mathscr{A}$ is a map $* : \mathscr{A} \to \mathscr{A}$ such that $(x^*)^* = x$ for all $x \in \mathscr{A}$. A *Banach $*$-algebra*[8] is a Banach algebra $\mathscr{A}$ with an involution operation $*$ that is

- *conjugate-linear*, i.e., $(x+y)^* = x^* + y^*$ and $(\lambda x)^* = \bar{\lambda} x^*$,
- *contravariant*, i.e., $(xy)^* = y^* x^*$, and

for all $x, y \in \mathscr{A}$ and $\lambda \in \mathbb{F}$. A Banach $*$-algebra $\mathscr{A}$ is *commutative* if $\mathscr{A}$ as a Banach algebra is commutative, and *unital* if $\mathscr{A}$ as a Banach algebra is unital and satisfies the additioanl identity

- $1^* = 1$.

To illustrate that the $*$-involution is indeed a generalization of the complex conjugate, we now recover the standard fact that $(A^{-1})^* = (A^*)^{-1}$ for all $A \in \mathscr{M}_n(\mathbb{F})$ in the Banach $*$-algebra setting.

Definition 16.48. An element x of a unital associative $\mathbb{F}$-algebra is *invertible* if there exists an element y such that $xy = yx = 1$, called an *inverse* of x. There is at most one inverse of x, and we write x^{-1} to denote the inverse if it exists.

Proposition 16.49. *An element x of a unital Banach $*$-algebra is invertible if and only if x^* is; if x is invertible, then $(x^*)^{-1} = (x^{-1})^*$.*

Proof. $1 = 1^* = (xx^{-1})^* = (x^{-1})^* x^*$ by contravariance. □

We now extend the notions of self-adjoint, normal, and unitary matrices to Banach $*$-algebras. As a particular instance of the following generalization, we obtain the definitions of self-adjoint, normal, and unitar operators.

Definition 16.50. An element of a Banach $*$-algebra $\mathscr{A}$ is *self-adjoint* if $x^* = x$ and *normal* if $x^* x = x x^*$. If $\mathscr{A}$ is unital, then $x \in \mathscr{A}$ is *unitary* if $x^* x = x x^* = 1$, or, equivalently, if $x^* = x^{-1}$.

We have seen in Proposition 11.18 that the adjoint operation preserves the operator norm. Since matrices of the form A^*A were of pivotal importance in the study of diagonalizable matrices in linear algebra, we seek to know more about operators of the form T^*T. An important basic property of such operators is the following proposition:

Proposition 16.51. $\|T^*T\| = \|T^*\|\|T\| = \|T\|^2$ *for all $T \in \mathscr{B}(\mathcal{H})$*

[8]Again, we could define a *$*$-algebra* by defining a $*$-involution operation on an associative $\mathbb{F}$-algebra, thereby dropping the "Banach" portion of the definition.

Proof. The adjoint operation is an isometry (Proposition 11.18), and so we have the identity $\|T^*\|\|T\| = \|T\|^2$. Since $\mathscr{B}(\mathcal{H})$ is a Banach algebra (Proposition 16.43), we have the norm inequality $\|T^*T\| \leq \|T^*\|\|T\|$, whence it suffices to establish the converse inequality. To this end, we fix a sequence $(x_n)_{n=1}^{\infty}$ in $\mathcal{H}$ such that $\|x_n\| = 1$ and $\lim_n \|Tx_n\| = \|T\|$ and observe that

$$\begin{aligned}
\|T^*T\| &= \sup_{\|x\|\leq 1} \|T^*Tx\| \\
\text{(def. of operator norm)} &\geq \lim_{n\to\infty} \|T^*Tx_n\| \\
(\|x_n\| = 1) &= \lim_{n\to\infty} \|T^*Tx_n\|\|x_n\| \\
\text{(Schwarz)} &\geq \lim_{n\to\infty} |\langle T^*Tx_n, x_n\rangle| \\
\text{(def. of Hilbert adjoint)} &= \lim_{n\to\infty} |\langle Tx_n, Tx_n\rangle| \\
\text{(def. of inner product norm)} &= \lim_{n\to\infty} \|Tx_n\|^2 \\
\text{(def. of } (x_n)) &= \|T\|^2 \\
\text{(isometry)} &= \|T^*\|\|T\|,
\end{aligned}$$

as was to be shown. □

The above property turns $\mathscr{B}(\mathcal{H})$ into a *unital C^*-algebra*, as advertised in the beginning of this section:

Definition 16.52. A Banach $*$-algebra $\mathscr{A}$ is a *C^*-algebra* if every element x of $\mathscr{A}$ satisfies the *C^*-identity*

$$\|x^*x\| = \|x\|^2.$$

A C^*-algebra $\mathscr{A}$ is *unital* or *commutative* if $\mathscr{A}$ as a Banach $*$-algebra is unital or commutative, respectively.

We observe that the identity akin to that in Proposition 16.51 holds for all C^*-algebras.

Proposition 16.53. *The $*$-involution on a C^*-algebra $\mathscr{A}$ is an isometry, whence*

$$\|x^*x\| = \|x^*\|\|x\| \tag{16.54}$$

for all $x \in \mathscr{A}$. Conversely, if the $*$-involution is an isometry, then (16.54) implies the C^*-identity.

Proof. If the C^* identity holds, then

$$\|x\|^2 = \|x^*x\| \leq \|x^*\|\|x\|,$$

and so $\|x^*\| \leq \|x\|$. Similarly,

$$\|x^*\|^2 = \|(x^*)^*x^*\| = \|xx^*\| \leq \|x\|\|x^*\|,$$

and so $\|x^*\| \geq \|x\|$. It follows that the $*$-involution is an isometry, whence (16.53) follows. The converse is trivially established. □

In fact, (16.54) implies (16.53) without the isometry condition on the $*$-involution
The theory of C^*-algebras provide powerful tools for studying operators. In fact, the abstract definition presented in Definition 16.52 grew out of the study of *operator algebras*, which we define below.

Definition 16.55. The *C^*-algebra of an operator $T \in \mathscr{B}(\mathcal{H})$*, denoted by $C^*(T)$,is the norm closure of the space of polynomials in two variables T and T^*. In other words, $C^*(T)$ is the norm closure of the span of the set

$$\{T^n\}_{n\in\mathbb{N}} \cup \{(T^*)^n\}_{n\in\mathbb{Z}}.$$

In general, a *C^*-algebra of operators* in $\mathscr{B}(\mathcal{H})$ is a closed linear subspace $\mathscr{A}$ of $\mathscr{B}(\mathcal{H})$ such that $T, U \in \mathscr{A}$ implies $TU \in \mathscr{A}$ and $T^* \in \mathscr{A}$.

Indeed, the name *C^*-algebra* comes from "Norm-**C**losed $*$-sub**algebra** of $\mathscr{B}(\mathcal{H})$." A fundamental theorm of Gelfand and Naimark shows that the study of C^*-algebras is no more general than the study of operator algebras.

Theorem 16.56 (Gelfand–Naimark theorem)**.** *Every C^*-algebra is isometrically $*$-isomorphic to a C^*-algebra of operators. In other words, for every C^*-algebra $\mathscr{A}$, there exists a Hilbert space $\mathcal{H}$, a norm-closed Banach $*$-subalgebra $\mathscr{B}$ of $\mathscr{B}(\mathcal{H})$, and an isometric isomorphism $\pi : \mathscr{A} \to \mathscr{B}$ such that $\pi(x^*) = \pi(x)^*$ for all $x \in \mathscr{A}$.*

See Chapter 4 of [Dou98], or Chapter 1 of [Arv76] for a proof. The former reference covers the topics of this chapter in much greater detail; the latter is a standard textbook in the basic theory of C^*-algebras, assuming some exposure to the theory of commutative Banach algebras.

The term *C^*-algebra* was coined by Irving Segal in [Seg47] and was used in the context of concrete operator algebras a là Definition 16.55. The abstract definition appeared as Definition 16.52 in this section was first introduced by Gelfand and Naimark [GN43] as a technical hypothesis for proving a weaker version of Theorem 16.56; here Banach algebras as in Definition 16.44 are referred to as *normed rings*, and Banach $*$-algebras (Definition 16.47) with isometric $*$-involutions as *$*$-rings*. Abstract C^*-algebras were first studied systematically by Charles Rickart in [Ric45], where he employs the term *B^*-algebras* for what we have named C^*-algebras in this section. But the full version of the Gelfand–Naimark theorem—which is the one we stated as Theorem 16.56—implies that C^*-algebras and B^*-algebras are precisely the same, and the division of concrete algebras as C^* and abstract algebras as B^* is no longer necessary. In fact, performing searches on widely used mathematical databases such as MathSciNet, MathOverflow, and Math StackExchange indicates that the term *B^*-algebras* is no longer in use, even though fairly modern textbooks such as [Lax02] and [Rud91] continue to use the term to refer to abstract C^*-algebras.

17. Appendix: Categories

In this appendix, we introduce a few elements of the category-theoretic language. This allows us to discuss the similarities between different mathematical structures in a clear, rigorous manner. We stress that the contents of this section are purely *linguistic*: we shall not attempt to discuss category theory proper, which can be found in, for example, [Mac98]. In particular, we shall not concern ourselves with foundational issues.

17.1. Basic Definitions.

Definition 17.1. A *category* $\mathcal{C}$ consists of the following:

- a class $\operatorname{Ob}\mathcal{C}$ of *objects*;
- a set $\operatorname{Hom}_{\mathcal{C}}(A, B)$ of *morphisms* for all $A, B \in \operatorname{Ob}\mathcal{C}$;

- an *identity morphism* $\mathrm{id}_A \in \mathrm{Hom}_{\mathcal{C}}(A, A)$ for each $A \in \mathrm{Ob}\,\mathcal{C}$;
- a *composition function* $\mathrm{Hom}_{\mathcal{C}}(A, B) \times \mathrm{Hom}_{\mathcal{C}}(B, C) \to \mathrm{Hom}_{\mathcal{C}}(A, C)$ whenever $A, B, C \in \mathrm{Ob}\,\mathcal{C}$.

We write $f : A \to B$ to denote a generic element of $\mathrm{Hom}_{\mathcal{C}}(A, B)$. Given $f : A \to B$ and $g : B \to C$, we write $g \circ f$ to denote the element of $\mathrm{Hom}_{\mathcal{C}}(A, C)$ produced by the composite function. The composition functions are subject to the following axioms:

- *Associativity.* $(h \circ g) \circ f = h \circ (g \circ f)$ whenever $f : A \to B$, $g : B \to C$, and $h : C \to D$;
- *Identity.* $\mathrm{id}_B \circ f = f \circ \mathrm{id}_A = f$ whenever $f : A \to B$.

A category $\mathcal{C}'$ is a *subcategory* of $\mathcal{C}$ if

- $\mathrm{Ob}\,\mathcal{C}'$ is a subclass of $\mathrm{Ob}\,\mathcal{C}$;
- for all $A, B \in \mathrm{Ob}\,\mathcal{C}'$, the set $\mathrm{Hom}_{\mathcal{C}'}(A, B)$ is a subset of $\mathrm{Hom}_{\mathcal{C}}(A, B)$;
- the morphisms of $\mathcal{C}'$ are subject to the same composition functions as those of $\mathcal{C}$.

Example 17.2. The simplest example is the category **Set** of sets. Here Ob(**Set**) is the class of all sets, and $\mathrm{Hom}_{\mathbf{Set}}(A, B)$ is simply the set of all functions from A to B. The category **FinSet** of finite sets with all functions as morphisms is a subcategory of **Set**.

Example 17.3. The category **Top** of topological spaces with continuous functions as morphisms is a category. The isomorphisms are homeomorphisms.

Example 17.4. The category $\mathbb{F}-$**Vect** of vector spaces over $\mathbb{F}$ with linear transformations as morphisms is a category. The isomorphisms are linear isomorphisms.

Example 17.5. Let us fix a base field $\mathbb{F}$. The category **TopVect** of topological vector spaces consists of topological vector spaces with continuous linear transformations as morphisms. f is an isomorphism in **TopVect** if and only if f is an isomorphism in **Top** and $\mathbb{F}-$**Vect**.

Example 17.6. The category **NLS** of normed linear spaces consists of normed linear spaces with bounded linear transformations as morphisms. The category **Ban** of Banach spaces is then a subcategory of **NLS**.

For technical reasons, category theorists often restrict the morphisms in **NLS** and **Ban** to be *short linear maps*, viz., linear maps $T : X \to Y$ whose norm $\|T\|$ is at most 1. One reason for this is that **NLS** and **Ban** as we have defined above have too many isomorphisms. Indeed, any bounded, bijective linear transformation whose inverse is bounded is an isomorphism in these categories, whether it is isometric (see Definition 8.11) or not. If we only consider short linear maps to be morphisms, then the categorical isomorphisms are surjective linear isometries.

To highlight this distinction, **NLS** and **Ban** with bounded linear transformations as morphisms are often referred to as *isomorphic categories*, and those with short linear maps as morphisms *isometric categories*.

Definition 17.7. Let $\mathcal{C}$ be a category, and let $A, B, C \in \mathrm{Ob}\,\mathcal{C}$. A morphism $f : A \to B$ is *monic*, or a *monomorphism*, if, for each pair of morphisms $g_1 : C \to A$ and $g_2 : C \to A$, the identity $f \circ g_1 = f \circ g_2$ implies that $g_1 = g_2$. $f : A \to B$ is *epi*, or an *epimorphism*, if, for each pair of morphisms $h_1 : B \to C$ and $h_2 : B \to C$, the identity $h_1 \circ f = h_2 \circ f$ implies that $h_1 = h_2$. $f : A \to B$ is an *isomorphism* if there exists a morphism $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$.

Example 17.8. In **Set**, monomorphisms are injective functions, epimorphisms are surjective functions, and isomorphisms are bijective functions.

Proof that monic = injective. Suppose that $f : A \to B$ is an injective function. Let $g_1 : C \to A$ and $g_2 : C \to B$ be functions on another set C such that $f \circ g_1 = f \circ g_2$. For each $x \in C$, we have $f(g_1(x)) = f(g_2(x))$, whence injectivity of f implies that $g_1(x) = g_2(x)$. We conclude that f is monic.

Conversely, we suppose that f is monic. We fix $x, y \in A$ and suppose that $f(x) = f(y)$. We let $C = \{0\}$ and define $g_1, g_2 : C \to A$ by setting $g_1(0) = x$ and $g_2(0) = y$. With these, we have the identity $f \circ g_1 = f \circ g_2$. Since f is monic, we conclude that $g_1 = g_2$, which implies that $x = y$. It follows that f is injective. □

Proof that epi = surjective. Suppose that $f : A \to B$ is a surjective function. Let $h_1 : B \to C$ and $h_2 : B \to C$ be functions into another set C such that $h_1 \circ f = h_2 \circ f$. This, in particular, implies that $h_1(x) = h_2(x)$ whenever $x \in \operatorname{im} f$. Since $\operatorname{im} f = B$, it follows that $h_1 = h_2$. We conclude that f is epi.

Conversely, we suppose that $f : A \to B$ is epi. Let $C = \{0, 1\}$ and define $h_1, h_2 : B \to C$ by setting $h_1(x) = 1$ for all $x \in B$ and

$$h_2(x) = \begin{cases} 1 & \text{if } x \in \operatorname{im} f; \\ 0 & \text{if } x \notin \operatorname{im} f. \end{cases}$$

By construction $h_1 \circ f = h_2 \circ f$, and so $h_1 = h_2$. This implies that $\operatorname{im} f = B$, and so f is surjective. □

It is trivial to show that isomorphisms in **Set** are precisely bijective functions. □

Proposition 17.9. *If $f : A \to B$ is an isomorphism, then there is precisely one morphism $g : B \to A$ such that $g \circ f = \mathrm{id}_A$ and $f \circ g = \mathrm{id}_B$. We say that g is the* inverse morphism *of f and denote it by f^{-1}.*

Proof. If g and h are two such morphisms, then

$$h = \mathrm{id}_A \circ h = (g \circ f) \circ h = g \circ (f \circ h) = g \circ \mathrm{id}_B = g,$$

as was to be shown. □

Definition 17.10. Let $\mathcal{C}$ be a category and $A \in \operatorname{Ob} \mathcal{C}$. An *endomorphism on A* is a morphism $f : A \to A$. If f is an isomorphism, then f is said to be an *automorphism on A*. The collection of automorphisms on A is denoted by $\operatorname{Aut}(A)$.

The following result is obvious.

Proposition 17.11. *If $f : A \to A$ and $g : A \to A$ are endomorphisms, then $g \circ f : A \to A$ is an endomorphism. If f and g are automorphisms, then so is $g \circ f$.* □

Remark 17.12. $\operatorname{Aut}(A)$ is always a *group* with respect to the function composition operation.

We shall use commutative diagrams as visual aids occasionally. The simplest diagram is of the form

$$A \xrightarrow{f} B,$$

which is just another way of writing $f : A \to B$. Multiple arrows denote a composite morphism. For example,

$$A \xrightarrow{f} B \xrightarrow{g} C$$

denotes the composite morphism $g \circ f : A \to C$ that we obtain from $f : A \to B$ and $g : B \to C$.

Multiple arrows to the same destination indicate that all possible paths should yield the same composite morphism. For example,

$$\begin{array}{ccc} A & \xrightarrow{f_1} & B \\ \downarrow{\scriptstyle f_2} & & \downarrow{\scriptstyle g_1} \\ C & \xrightarrow{g_2} & D \end{array}$$

means the two composite morphisms

$$A \xrightarrow{f_1} B \xrightarrow{g_1} D$$

and

$$A \xrightarrow{f_2} C \xrightarrow{g_2} D$$

are identical. In this case, we say that the diagram *commutes*.

The arrows can be marked to specify various properties of morphisms. For example, we write

$$A \overset{f}{\hookrightarrow} B$$

to denote a monomorphism $f : A \to B$, and

$$A \overset{g}{\twoheadrightarrow} B$$

to denote an epimorphism $g : A \to B$.

$$A \overset{h}{\dashrightarrow} B$$

indicates that h is the unique morphism satisfying some properties.

17.2. **Universality.** A fundamental notion in the category-theoretic way of thinking is that of universality. Instead of giving a formal definition of universality, we introduce two universal constructions that will occur throughout these notes.

Definition 17.13. Let $\mathcal{C}$ be a category and $\{A_i\}_{i\in I}$ a set of objects in $\mathcal{C}$. The pair $(P, \{p_i\}_{i\in I})$ of an object P in $\mathcal{C}$ and a set $\{p_i\}_{i\in I}$ of morphismis $p_i : P \to A_i$ is said to be a *product* of $\{A_i\}_{i\in I}$ if every pair $(Q, \{q_i\}_{i\in I})$ of an object Q in $\mathcal{C}$ and a set $\{q_i\}_{i\in I}$ of morphisms $q_i : Q \to A_i$ admits a unique morphism $q : Q \to P$ such that the diagram

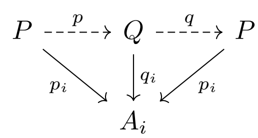
$$\begin{array}{ccc} Q & \overset{q}{\dashrightarrow} & P \\ & {\scriptstyle q_i}\searrow & \downarrow{\scriptstyle p_i} \\ & & A_i \end{array}$$

commutes for all $i \in I$.

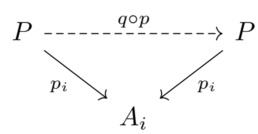
Example 17.14. In **Set**, the cartesian product $\prod_i X_i$ of sets $\{X_i\}$ with the usual coordinate functions $\pi_i : \prod_i X_i \to X_i$ is a product. Indeed if we have a pair $(Y, \{f_i\}_{i\in I})$ of a set Y and a collection $\{f_i\}_{i\in I}$ of functions $f_i : Y \to X_i$, then the function $f : Y \to \prod_i X_i$ that sends y to $(f_i(y))_{i\in I}$ is evidently the unique function such that $\pi_i \circ f = f_i$ for all $i \in I$.

Proposition 17.15. *Let $\mathcal{C}$ be a category, and $\{A_i\}_{i\in I}$ a set of objects in $\mathcal{C}$. If $(P, \{p_i\}_{i\in I})$ and $(Q, \{q_i\}_{i\in I})$ are products of $\{A_i\}_{i\in I}$, then there exists an isomorphism from P to Q. We write $(\prod_{i\in I} A_i, \{p_i\}_{i\in I})$ to denote any one of these isomorphic product objects; the morphisms p_i are referred to as the* canonical projection maps.

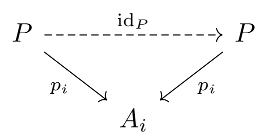
Proof. By the definition of products, there exist morphisms $p : P \to Q$ and $q : Q \to P$ such that the diagram



commutes for all $i \in I$. This implies that the diagram

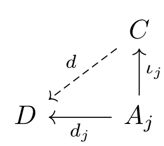


commutes for all $i \in I$. Since



also commutes for all $i \in I$, the uniqueness clause in the definition of products implies that $q \circ p = \mathrm{id}_P$. Similarly, we can show that $p \circ q = \mathrm{id}_Q$, whence $p : P \to Q$ is an isomorphism. □

Definition 17.16. Let $\mathcal{C}$ be a category and $\{A_j\}_{j\in I}$ a set of objects in $\mathcal{C}$. The pair $(C, \{\iota_j\}_{j\in I})$ of an object C in $\mathcal{C}$ and a set $\{\iota_j\}_{j\in I}$ of morphismis $\iota_j : A_j \to C$ is said to be a *coproduct* of $\{A_i\}_{j\in I}$ if every pair $(D, \{d_j\}_{j\in I})$ of an object D in $\mathcal{C}$ and a set $\{d_j\}_{j\in I}$ of morphisms $d_j : A_j \to D$ admits a unique morphism $d : C \to D$ such that the diagram



commutes for all $j \in I$.

Example 17.17. In **Set**, the *disjoint union*

$$\coprod_j X_j = \bigcup_{j\in I} \{(x_j, j) : x_j \in X_j\}$$

with the usual injection maps $\iota_j(x) = (x, j)$ is a coproduct.

Proposition 17.18. *Let $\mathcal{C}$ be a category, and $\{A_j\}_{j\in I}$ a set of objects in $\mathcal{C}$. If $(C, \{\iota_j\}_{j\in I})$ and $(D, \{d_j\}_{j\in I})$ are products of $\{A_j\}_{j\in I}$, then there exists an isomorphism from C to D. We write $(\coprod_{j\in I} A_j, \{\iota_j\})$ to denote any one of these isomorphic product objects; the morphisms ι_j are referred to as the* canonical injection maps.

Proof. Proceed as in the proof of Proposition 17.15. □

17.3. **Natural Transformations.** A *covariant functor* $F : \mathcal{C} \to \mathcal{D}$ from a category $\mathcal{C}$ to another category $\mathcal{D}$ is a rule that sends each object $A \in \operatorname{Ob}\mathcal{C}$ to $F(A) \in \operatorname{Ob}\mathcal{D}$ and every morphism $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$ to $F(f) \in \operatorname{Hom}_{\mathcal{D}}(F(A), F(B))$. A *contravariant functor* $G : \mathcal{C} \to \mathcal{D}$ from $\mathcal{C}$ to $\mathcal{D}$ is a rule that sends each object $A \in \operatorname{Ob}\mathcal{C}$ to $G(A) \in \operatorname{Ob}\mathcal{D}$ and every morphism $f \in \operatorname{Hom}_{\mathcal{C}}(A, B)$ to $G(f) \in \operatorname{Hom}_{\mathcal{D}}(G(B), G(A))$.

Example 17.19. Define the *dual functor* $D : \mathbf{NLS} \to \mathbf{NLS}$ by setting $D(V) = V^*$ and $D(T) = T^*$, the topological adjoint of T (Definition 11.17). This is a contravariant functor, as the "arrow" $V \xrightarrow{T} W$ reverses its direction when the adjoint is taken: $W^* \xrightarrow{T^*} V^*$.

The *double dual functor* D^2 on **NLS**. D^2 is given by the composition of the dual functor D D^2 is a covariant functor, as the arrow $V \xrightarrow{T} W$ gets flipped twice. For illustrative purposes, we write down the action of $T^{**} : V^{**} \to W^{**}$ explicitly: for each $\varphi \in V^{**}$ and every $l \in W^*$, we have

$$(T^{**}\varphi)l = (\varphi T^*)l = \varphi(T^* l) = \varphi(lT).$$

Example 17.20. The dual functor is a special case of *Hom functors*. Given a category $\mathcal{C}$ and an object A in $\mathcal{C}$, we define the *covariant Hom functor* $\operatorname{Hom}(A, -)$ from $\mathcal{C}$ to **Set** by sending each $B \in \operatorname{Ob}\mathcal{C}$ to $\operatorname{Hom}_{\mathcal{C}}(A, B)$ and $f \in \operatorname{Hom}_{\mathcal{C}}(B, C)$ to $f^* = \operatorname{Hom}(A, f) \in \operatorname{Hom}_{\mathbf{Set}}(\operatorname{Hom}_{\mathcal{C}}(A, B), \operatorname{Hom}_{\mathcal{C}}(A, C))$, given by the formula

$$f^*(g) = f \circ g.$$

Similarly, we can define the *contravariant Hom functor* $\operatorname{Hom}(-, A)$ from $\mathcal{C}$ to **Set** by sending each $B \in \operatorname{Ob}\mathcal{C}$ to $\operatorname{Hom}_{\mathcal{C}}(B, A)$ and $f \in \operatorname{Hom}_{\mathcal{C}}(B, C)$ to $f_* = \operatorname{Hom}(f, A) \in \operatorname{Hom}\mathbf{Set}(\operatorname{Hom}_{\mathcal{C}}(C, A), \operatorname{Hom}_{\mathcal{C}}(B, A))$, given by the formula

$$f_*(g) = g \circ f.$$

In this sense, the dual functor D can be thought of as the contravariant Hom functor $\operatorname{Hom}_{\mathbf{NLS}}(-, \mathbb{F})$. Here the base field $\mathbb{F}$ is considered as a one-dimensional normed linear space with some norm. Note that the Hom-set $\operatorname{Hom}_{\mathbf{NLS}}(X, \mathbb{F})$ is precisely the dual space X^*, which, in fact, is given the structure of a normed linear space. Therefore, the contravariant Hom functor $\operatorname{Hom}_{\mathbf{NLS}}(-, \mathbb{F})$ also maps **NLS** into **NLS**.

Given two covariant functors F and G from $\mathcal{C}$ to $\mathcal{D}$, we can define a *natural transformation (of covariant functors)* $\alpha : F \to G$ from F to G as a rule that associates a morphism $\alpha_A \in \operatorname{Hom}_{\mathcal{D}}(F(A), G(A))$ to each object $A \in \operatorname{Ob}\mathcal{C}$ such that the diagram

$$\begin{array}{ccc} F(A) & \xrightarrow{Ff} & F(B) \\ {\scriptstyle \alpha_A}\downarrow & & \downarrow{\scriptstyle \alpha_B} \\ G(A) & \xrightarrow[Gf]{} & G(B) \end{array}$$

commutes, i.e., $(Gf)\alpha_A = \alpha_B(Ff)$, for all $A, B \in \operatorname{Ob}\mathcal{C}$. If each α_A is an isomorphism, then α is said to be a *natural isomorphism.*

Example 17.21. As an example, we establish a natural transformation from the identity functor I that does nothing to the objects and the morphisms to the double dual functor D^2. Recall that there is a canonical embedding $\alpha_V : V \to V^{**}$ of a vector space V into its double dual V^{**}, which associates each $x \in V$ to $\hat{x} \in V^{**}$. $\hat{x}$ is given by the formula

$$\hat{x}l = lx$$

for each $l \in V^*$. We claim that $\alpha : I \to D^2$ that associates V to α_V is a natural transformation from the identity functor I to the double dual functor D^2. In other words, the diagram

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ {\scriptstyle \alpha_V}\downarrow & & \downarrow{\scriptstyle \alpha_W} \\ V^{**} & \xrightarrow[T^{**}]{} & W^{**} \end{array}$$

commutes for all normed linear spaces V and W. To prove this assertion, we first note that $T^{**}\alpha_V x$ for a fixed $x \in V$ is an element of W^{**}, hence it takes elements of W^* and produces a scalar. We now fix $l \in W^*$ and observe that

$$\begin{aligned} (T^{**}\alpha_V x)l &= (T^{**}\hat{x})l = (\hat{x}T^*)l = \hat{x}(T^*l) = \hat{x}(lT) \\ &= (lT)x = l(Tx) = (\widehat{Tx})l = (\alpha_W Tx)l, \end{aligned}$$

which is the desired result.

Example 17.22. We remark that α_V in the above example is an isomorphism if V is finite-dimensional, hence α is a natural isomorphism if we restrict to the category of finite-dimensional normed linear spaces. The same argument shows that α is a natural transformation on the category of Banach spaces, and a natural isomorphism on the category of reflexive Banach spaces.

18. Appendix: Sets

Formally, a *set* is any object that satisfies the axioms of *Zermelo–Fraenkel set theory with the Axiom of Choice* (ZFC). For our purposes, however, the exact details are immaterial, and it suffices to think of a set as a collection X such that every mathematical object x should either be an element of X or not, but not both. We take for granted that the reader is acquainted with basic notions of set theory such as relations, functions, and orders, as well as many examples of sets, including

- the empty set $\varnothing$,
- the set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$,
- the set of positive integers numbers $\mathbb{Z}^+ = \{1, 2, \ldots\}$,
- the set of integers $\mathbb{Z} = \{\ldots, -3, -2, 1, 0, 1, 2, 3, \ldots\}$,
- the set of rational numbers $\mathbb{Q}$,
- the set of real numbers $\mathbb{R}$, and
- the set of complex numbers $\mathbb{C}$.

The purpose of this section is merely to spell out the nontrivial set-theoretic results that we shall call upon in these notes. While the primary references for this section are Chapters 2 and 3 of [Jec06], [HJ99] is recommended for readers unfamiliar with basic set theory. The latter contains detailed exposition at a more leisurely pace than the former.

18.1. **Order theory.** Among the main objects of study in set theory are posets, or partially ordered sets.

Definition 18.1. A *partial order* on a set X is a relation $\preceq$ on X that satisfies the following properties:

(1) **reflexivity**. $x \preceq x$ for all $x \in X$;
(2) **antisymmetry**. if $x \preceq y$ and $y \preceq x$, then $x = y$;
(3) **transitivity**. if $x \preceq y$ and $y \preceq z$, then $x \preceq z$.

A *partialled ordered set*, or a *poset* is a pair $(X, \preceq)$ of a set X equipped with a partial order $\preceq$.

Example 18.2. The usual less-than-or-equal-to relations on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are partial orders. □

Example 18.3. Recall that the *power set* $\mathcal{P}(X)$ of a set X is the collection of all subsets of X. Given a set X, the $\subseteq$-relation on $\mathcal{P}(X)$ is a partial order. □

We take the category-theoretic approach and introduce at once the notion of structure-preserving maps between posets.

Definition 18.4. An *order-preserving map*, or a *morphism of posets* between two posets $(X, \preceq)$ and $(Y, \leq)$ is a function $f : X \to Y$ such that $x_1 \preceq x_2$ implies $f(x_1) \leq f(x_2)$.

Proposition 18.5. *The collection of all posets with order-preserving maps as morphisms forms a category, denoted by* **poset**. *Moreover:*

(1) monomorphisms in **poset** *are injective order-preserving maps;*
(2) epimorphisms in **poset** *are surjective order-preserving maps;*
(3) isomorphisms in **poset**, *called* order isomorphisms, *are bijective order-preserving maps whose inverses are also order-preserving.*

Proof. Checking the category axioms is trivial.

(1) Let $f : (X, \preceq) \to (Y, \leq)$ be a function. Suppose that f is a monomorphism in **poset**. This, in particular, implies that f is a monomorphism in **Set**, whence Example 17.8 implies that f is an injective function. The assumption also implies that f is an order-preserving map.

Conversely, suppose that f is an injective order-preserving map. Let $g_1, g_2 : (Z, \lesssim) \to (X, \preceq)$ be order-preserving maps such that $f \circ g_1 = f \circ g_2$. This, in particular, implies that $f \circ g_1 = f \circ g_2$ in **Set**, and so Example 17.8 implies that $g_1 = g_2$.

(2) Let $f : (X, \preceq) \to (Y, \leq)$ be a function. Suppose that f is an epimorphism in **poset**. This, in particular, implies that f is an epimorphism in **Set**, whence Example 17.8 implies that f is a surjective function. The assumption also implies that f is an order-preserving map.

Conversely, suppose that f is a surjective order-preserving map. Let $h_1, h_2 : (Y, \leq) \to (Z, \lesssim)$ be order-preserving maps such that $h_1 \circ f = h_2 \circ f$. This, in particular, implies that $h_1 \circ f = h_2 \circ f$ in **Set**, and so Example 17.8 implies that $h_1 = h_2$.

(3) This is trivial. □

Example 18.6. To qualify as an isomorphism in **poset**, it is not enough for an order-preserving map to be merely bijective. Indeed, we let X be a nonempty set,

$\preceq$ be a partial order on X, and $\leq$ be $X \times X$, so that $x \leq y$ for all $x, y \in X$. Then every bijection $f : X \to X$ is order-preserving, but it is not hard to pick $\preceq$ and f such that f^{-1} is not order-preserving. $\square$

It is often useful to consider the following variant of partial orders:

Definition 18.7. A *strict partial order* on a nonempty set X is a relation $\prec$ on X that satisfies the following properties:

(1) **irreflexivity**. $x \not\prec$ for all $x \in X$;
(2) **transitivity**. if $x \prec y$ and $y \prec z$, then $x \prec z$.

We can define order-preserving maps for strict partial orders analogously.

Remark 18.8. If $\preceq$ is a partial order on X, then we can define a strict partial order $\prec$ on X by declaring $x \prec y$ if and only if $x \preceq y$ and $x \neq y$. Similarly, if $\prec$ is a strict partial order on X, then we can define a partial order $\preceq$ on X by declaring $x \preceq y$ if and only if either $x \prec y$ or $x = y$.

Order-theoretic concepts such as maximum, minimum, supremum, and infimum admit straightforward generalizations in the context of partial orders:

Definition 18.9. Let X be a nonempty set, $\preceq$ be a partial order on X, and E be a subset of X.

(1) An *upper bound of* E *in* X is an element $u \in X$ such that $u \succeq x$ for all $x \in E$. A *supremum of* E *in* X is an upper bound s of E in X such that $s \preceq u$ whenever u is another upper bound of E in X; $\sup E$ is necessarily unique.
(2) A *lower bound of* E *in* X is an element $l \in X$ such that $l \preceq x$ for all $x \in E$. An *infimum of* E *in* X is a lower bound i of E in X such that $i \succeq l$ whenever l is another lower bound of E in X; $\inf E$ is necessarily unique.
(3) A *greatest element of* E is an upper bound of E in X that is also an element of E,
(4) A *least element of* E is a lower bound of E that is also an element of E.
(5) A *maximal element of* E is an element $m \in E$ such that no $x \in E$ satisfies the relation $x \succ m$,
(6) A *minimal element of* E is an element $m \in E$ such that no $x \in E$ satisfies the relation $x \prec m$.

The usual orders on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ are more than just partial orders. We abstract the additional properties of these order relations:

Definition 18.10. A (strict) partial order $\preceq$ on X is a *(strict) linear order*, or a *(strict) total order* if $\leq$ satisfies the *trichotomy principle*, i.e., one of the following three must hold for arbitrary $x, y \in X$:

(1) $x \prec y$;
(2) $x = y$;
(3) $x \succ y$.

Definition 18.11. A linear order $\prec$ on a nonempty set X is a *well-order* if each nonempty subset E of X contains an element $m \in E$ such that $m \preceq x$ for all $x \in E$.

Example 18.12. $\mathbb{N}$ with $<$ is a well-ordered set.

Example 18.13. $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$ with $<$ strictly linearly ordered sets but not well-ordered sets. Indeed, $\{-n : n \in \mathbb{N}\}$ is a subset of each of these sets that is not bounded below.

Our goal is to devise a generalization of the natural number system such that every well-ordered set is order-isomorphic to one of the numbers in this generalized system. The crucial property is as follows:

Definition 18.14. A set X is *transitive* if every element of X is a subset of X.

Example 18.15 (Von Neumann definition of natural numbers)*.* A constructive way of defining the natural numbers in ZFC is as follows.

- $0 = \varnothing$;
- $1 = \{0\} = \{\varnothing\}$;
- $2 = \{0, 1\} = \{\varnothing, \{\varnothing\}\}$;
- $\cdots$
- $n = \{0, 1, \ldots, n-1\}$;
- $\cdots$

Note that, under this construction, each natural number is a transitive set. Moreover, the element relation $\in$ is a strict well-order on each natural number. Indeed, $n < m$ in the usual sense if and only if $n \in m$ in our construction, and the axioms of a well-order can be checked easily. □

In light of the above construction, we introduce the following definition:

Definition 18.16. An *ordinal number*, or an *ordinal* is a set that is transitive and well-ordered by the element relation $\in$. Given two ordinals α and β, we write $\alpha < \beta$ if and only if $\alpha \in \beta$.

It can be proved that $\alpha \cup \{\alpha\}$ is an ordinal whenever α is an ordinal. In light of this, we write $\alpha + 1$ to denote $\alpha \cup \{\alpha\}$ and make the following definition:

Definition 18.17. An ordinal α is a *successor ordinal* if there exists an ordinal β such that $\alpha = \beta + 1$; if not, then α is said to be a *limit ordinal*.

Example 18.18. As per Example 18.15, every natural number other than 0 is a successor ordinal. The set of all natural numbers $\omega = \mathbb{N}$ is a limit ordinal, called the *first infinite ordinal*. □

The key property of ordinal numbers is the following infinitary generalization of mathematical induction:

Theorem 18.19 (Transfinite induction)**.** *Let $\mathcal{C}$ be a collection of ordinals. If $\mathcal{C}$ satisfies the following properties, then $\mathcal{C}$ contains all ordinals:*

(1) $0 \in \mathcal{C}$*;*
(2) if $\alpha \in \mathcal{C}$*, then* $\alpha + 1 \in \mathcal{C}$*;*
(3) if α *is a nonzero limit ordinal and* $\beta \in \mathcal{C}$ *for all ordinals* $\beta < \alpha$*, then* $\alpha \in \mathcal{C}$*.*

Proof. We shall make use of the following fact, whose proof we omit:

Claim. *If $\mathcal{C}$ is a nonempty collection of ordinals, then* $\inf \mathcal{C}$ *is an ordinal.*

Suppose for a contradiction that $\mathcal{C}$ does not contain all ordinal numbers. Let α be the least such ordinal, which exists by the claim. If α is a successor ordinal, then we can find $\beta \in \alpha$ such that $\beta + 1 = \alpha$. Since α is the least ordinal in $\mathcal{C}$, β must be in $\mathcal{C}$, whence (2) implies that $\alpha \in \mathcal{C}$. If α is a nonzero limit ordinal, then (3) and the least-ordinal property imply that $\alpha \in \mathcal{C}$. Both cases reach a contradiction, and so we conclude that $\mathcal{C}$ contains all ordinals. □

An important consequence of transfinite induction is *Zorn's lemma*, which is precisely the incarnation of infinitary mathematical induction that is used in most fields of mathematics. To this end, we shall need a preliminary definition.

Definition 18.20. Let X be a nonempty set and $\preceq$ a partial order on X. A *chain* in X is a subset E of X such that the restriction of $\preceq$ on E is a linear order.

Theorem 18.21 (Zorn's lemma)**.** *Let X be a nonempty set with a partial order $\preceq$. If every chain in X has an upper bound in X, then X has a maximal element.*

Proof. We let $c : \mathcal{P}(X) \smallsetminus \{\varnothing\} \to \bigcup(\mathcal{P}(X) \smallsetminus \{\varnothing\})$ be a *choice function*, i.e., a function that picks out an element of each of the sets in the collection $\mathcal{P}(X) \smallsetminus \{\varnothing\}$.

We let $a_0 = c(X)$ and define a_α inductively for each ordinal α by declaring a_α to be an element of X such that $a_\alpha \succ \alpha_\beta$ whenever $\beta < \alpha$; we terminate the process when there is no such element a_α. To check that this process continues until the point of termination, we must check the limit ordinal stages, as per the principle of transfinite induction (Theorem 18.19). But indeed, if α is a limit ordinal, then $\{a_\beta : \beta < \alpha\}$ is a chain in X, whence by assumption we can find an upper bound a_α.

We can now find an ordinal γ such that no $x \in X$ satisfies the relation $x > a_\gamma$. It follows that a_γ is a maximal element of X. □

Remark 18.22. We have tacitly made use of the *axiom of choice*, which guarantees the existence of choice functions. Moreover, the definition of $(a_\alpha)_\alpha$ requires *transfinite recursion*, which, essentially, allows us to make infinitary recursive definitions.

18.2. **Cardinality.** Let us now examine the problem of comparing sizes of sets. Two sets X and Y are said to be *equinumerous*, or *of the same cardinality*, if there exists a bijection $f : X \to Y$. We write $|X| = |Y|$ to denote that X and Y are equinumerous. In case f is merely an injection, we say that *the cardinality of X is at most the cardinality of Y* and write $|X| \leq |Y|$.

Example 18.23. Since

$$f(n) = \begin{cases} 2n & \text{if } n \geq 0; \\ 2(-n) - 1 & \text{if } n < 0 \end{cases}$$

is a bijection from $\mathbb{Z}$ to $\mathbb{N}$, we conclude that $|\mathbb{N}| = |\mathbb{Z}|$. □

Two basic properties of the "less than or equal to" cardinality relation are the following.

Proposition 18.24 (Partition principle)**.** $|X| \leq |Y|$ *if and only if there exists a surjection $g : Y \to X$.*

Proof. If there exists an injection $f : X \to Y$, then the codomain restriction $f' : X \to \operatorname{im} f$ is a bijection. We define a surjection $g : Y \to X$ as follows: fix $x_0 \in X$ and set

$$g(y) = \begin{cases} (f')^{-1}(y) & \text{if } y \in \operatorname{im} f; \\ x_0 & \text{if } y \notin f. \end{cases}$$

Conversely, if $g : Y \to X$ is a surjection, then the set $P = \{g^{-1}(\{x\}) : x \in X\}$ is a partition of Y. The function $f : X \to Y$ that sends each $x \in X$ to some element of $g^{-1}(\{x\})$ is an injection. □

Theorem 18.25 (Cantor–Bernstein)**.** *If* $|X| \leq |Y|$ *and* $|Y| \leq |X|$*, then* $|X| = |Y|$*.*

Proof. We shall make use of the following claim:

Claim. *If* $A_1 \subseteq B \subseteq A$ *and* $|A_1| = |A|$*, then* $|A_1| = |B| = |A|$*.*

We assume the claim for now and suppose that there exist injections $i : X \to Y$ and $j : Y \to X$. This, in particular, implies that the composite function $j \circ i : X \to X$ is an injective function, and so $|X| = |\operatorname{im}(j \circ i)|$. Now, $\operatorname{im}(j \circ i) \subseteq \operatorname{im} j|_{\operatorname{im} i} \subseteq \operatorname{im} j$. Since $\operatorname{im} j \subseteq X$, the claim implies that

$$|X| = |\operatorname{im}(j \circ i)| = |\operatorname{im} j|.$$

Since $\operatorname{im} j = Y$, we conclude that $|X| = |Y|$.

It therefore suffices to prove the claim. To this end, we take a bijection $\varphi : A \to A_1$ and define three collections of sets $\{A_n\}_{n \in \mathbb{N}}$, $\{B_n\}_{n \in \mathbb{N}}$, and $\{C_n\}_{n \in \mathbb{N}}$ as follows:

(1) let $A_0 = A$ and $B_0 = B$;
(2) for each $n \geq 2$, we define $A_n = \varphi(A_{n-1})$;
(3) for each $n \geq 1$, we define $B_n = \varphi(B_{n-1})$;
(4) for each $n \in \mathbb{N}$, we define $C_n = A_n \smallsetminus B_n$.

We also let

$$C = \bigcup_{n=0}^{\infty} C_n.$$

By definition, $A_0 \supseteq A_1$. Since $A_2 = \varphi(A_1)$ must be a subset of $\operatorname{im} \varphi = A_1$, we see that $A_1 \supseteq A_2$. We now fix $N \geq 3$ and assume inductively that $A_{n-1} \supseteq A_n$ for all $n < N$. Since $A_{N-2} \supseteq A_{N-1}$, we see that

$$A_{N-1} = \varphi(A_{N-2}) \supseteq \varphi(A_{N-1}) = A_N.$$

It follows that $A_{k-1} \supseteq A_k$ for all $k \in \mathbb{N}$. Similarly, $B_{k-1} \supseteq B_k$ for all $k \in \mathbb{N}$.

Since $A_0 \supseteq B_0$, we see that

$$A_1 = \varphi(A_0) \supseteq \varphi(B_0) = B_1.$$

Repeating this process, we conclude that $A_k \supseteq B_k$ for all $k \in \mathbb{N}$.

We now claim that $\varphi(C_k) = C_{k+1}$ for all $k \in \mathbb{N}$. To see this, we fix $k \in \mathbb{N}$ and pick an arbitrary $x \in C_k$. Since $C_k \subseteq A_k$, we see that

$$\varphi(x) \in \varphi(C_k) \subseteq \varphi(A_k) = A_{k+1}.$$

Moreover, $x \notin B_k$, and so injectivity of φ implies that $\varphi(x) \notin \varphi(B_k) = B_{k+1}$. It follows that $x \in A_{k+1} \smallsetminus B_{k+1} = C_{k+1}$. Since x was arbitrary, we conclude that $\varphi(C_k) \subseteq C_{k+1}$.

To show that $C_{k+1} = \varphi(C_k)$, we fix an arbitrary $y \in C_{k+1}$. Since $y \in C_{k+1} \subseteq A_{k+1}$, relation $\varphi(A_k) = A_{k+1}$ implies that we can find an $x \in A_k$ such that $\varphi(x) = y$. Now, if $x \in B_k$, then $\varphi(x) \in \varphi(B_k) = B_{k+1}$, which would, in turn, imply that $\varphi(x) \notin C_{k+1}$. This is evidently absurd, and so $x \notin B_k$. We conclude that $x \in C_k$, so that $y \in \varphi(C_k)$. Since y was arbitrary, the desired result now follows.

An important consequence of relation $C_{k+1} = \varphi(C_k)$ is that

$$\varphi(C) = \bigcup_{k=1}^{\infty} C_k.$$

We set $D = A \setminus C$ and define

$$f(x) = \begin{cases} \varphi(x) & \text{if } x \in C; \\ x & \text{if } x \in D. \end{cases}$$

$f|_C$ is precisely a restriction of φ and so it is injective. Similarly, $f|_D$ is a restriction of the identity function , and so it is injective. It follows that f is injective. Moreover,

$$f(C) \cup f(D) = \left(\bigcup_{n=1}^{\infty} A_n \setminus B_n \right) \cup \left[A \setminus \left(\bigcup_{n=0}^{\infty} A_n \setminus B_n \right) \right] = B,$$

and so f is surjective. It follows that f is a bijection from A to B, whence $|A| = |B|$. The claim is now established. □

We shall construct a generalization of natural numbers that allows us to quantify sizes of infinite sets. We first show that such an endeavor is nontrivial by exhibiting many different orders of infinity:

Theorem 18.26 (Cantor)**.** $|X| < |\mathcal{P}(X)|$ *for every set* X*.*

Proof. The function $f(x) = \{x\}$ is injection from X to $\mathcal{P}(X)$, and so $|X| \leq |\mathcal{P}(X)|$. Now, if $g : X \to \mathcal{P}(X)$ is an arbitrary function, then

$$D_g = \{x \in X : x \notin g(x)\}$$

is not in the subset of $\operatorname{im} g$. Therefore, g is not a surjection, and so $|X| \neq |\mathcal{P}(X)|$. □

Example 18.27. Every real number admits a unique binary expansion, and so $|\mathbb{R}| = |\mathcal{P}(\mathbb{Z})|$. By Cantor's theorem,

$$|\mathbb{N}| = |\mathbb{Z}| < |\mathcal{P}(\mathbb{Z})| = |\mathbb{R}|,$$

and so $\mathbb{R}$ is uncountable.

It cannot be determined within ZFC whether there exists a set X such that

$$|\mathbb{N}| < |X| < |\mathbb{R}|;$$

this is the *continuum hypothesis.* □

Since it can be proved that every well-ordered set is order-isomorphic to a unique ordinal, we might quantify the size of each well-ordered set by its isomorphic ordinal. The problem is that different ordinals can be of the same cardinality. Indeed, $|\omega| = |\omega + 1|$, as the mapping

$$f(\alpha) = \begin{cases} \alpha + 1 & \text{if } \alpha < \omega \\ 0 & \text{if } \alpha = \omega \end{cases}$$

is a bijection from $\omega + 1$ to ω. We therefore make the following definition:

Definition 18.28. A *cardinal number* is an ordinal α such that $|\alpha| \neq |\beta|$ whenever $\beta < \alpha$.

It now follows at once that every well-ordered set X admits a unique cardinal number κ such that $|X| = \kappa$. By Cantor's theorem (Theorem 18.26), every cardinal number admits a cardinal number greater than itself. This leads us to introduce the following enumeration of infinite cardinal numbers:

$$\aleph_0, \aleph_1, \dots, \aleph_n, \dots, \aleph_\omega, \aleph_{\omega+1}, \dots.$$

Here $\aleph_0 = \omega$. $\aleph_1$ as an ordinal number is called the *first uncountable ordinal* and is denoted by ω_1.

The restriction that the set needs to be well-ordered in order to assign a cardinal number is immaterial:

Theorem 18.29 (Well-ordering theorem)**.** *Every set can be well-ordered.*

Proof. Let X be a set. If X is finite, then well-ordering can be done by hand. We therefore assume that X is an infinite set.

We make use of the following fact, whose proof we omit.

Claim. *Ordinals are well-ordered by $<$.*

We use transfinite induction to label each element of the set X by a unique ordinal, whence X can be well-ordered by the claim. We let $c : \mathcal{P}(X) \setminus \{\varnothing\} \to \bigcup(\mathcal{P}(X) \setminus \{\varnothing\})$ be a choice function (see the proof of Theorem 18.21). We set $a_0 = c(X)$ and define a_α inductively for each ordinal α by setting

$$a_\alpha = c(X \setminus \{a_\beta : \beta < \alpha\});$$

we terminate the process when $X \setminus \{a_\beta : \beta < \alpha\}$ is empty.

We now let γ be the least ordinal number such that $X = \{a_\beta : \beta < \gamma\}$. The set $\{\alpha_\beta : \beta < k\}$ consists precisely of the elements of X, labeled by ordinals up to γ. We define an order $\preceq$ on X by declaring $a_\alpha \preceq a_\beta$ if and only if $\alpha < \beta$ or $\alpha = \beta$. It follows from the well-ordering of ordinals that $\preceq$ is a well-order on X. □

Remark 18.30. Once again, we have made use of the axiom of choice and transfinite recursion in the above proof.

We can define arithmetic operations on cardinal numbers; here A^B is the B-fold cartesian product of A, i.e.,

$$\prod_{x \in B} A.$$

Definition 18.31 (Cardinal arithmetic)**.** Given two cardinals κ and λ, we define the arithmetic operations as follows:

(1) $\kappa + \lambda = |A \cup B|$, where A and B are disjoint sets such that $|A| = \kappa$ and $|B| = \lambda$;
(2) $\kappa \cdot \lambda = |A \times B|$, where $|A| = \kappa$ and $|B| = \lambda$;
(3) $\kappa^\lambda = |A^B|$, where $|A| = \kappa$ and $|B| = \lambda$.

It can be shown that these operations are well-defined, i.e., they are independent of the choice of A and B.

Here are some basic facts about cardinal arithmetic, whose proofs are easy:

Proposition 18.32. *Let κ, λ, and μ be cardinal numbers.*

(1) $\kappa + \lambda = \lambda + \kappa$;
(2) $\kappa \cdot \lambda = \lambda \cdot \kappa$;
(3) $(\kappa + \lambda) \cdot \mu = \kappa \cdot \mu + \lambda \cdot \mu$;
(4) $(\kappa \cdot \lambda)^\mu = \kappa^\mu \cdot \lambda^\mu$;
(5) $\kappa^{\lambda + \mu} = \kappa^\lambda \cdot \kappa^\mu$;

(6) $(\kappa^\lambda)^\mu = \kappa^{\lambda \cdot \mu}$*;*
(7) If $\kappa \leq \lambda$*, then* $\kappa^\mu \leq \lambda^\mu$*;*
(8) If $0 < \lambda \leq \mu$*, then* $\kappa^\lambda \leq \kappa^\mu$*;*
(9) $\kappa^0 = 1$*;*
(10) $1^\kappa = 1$*;*
(11) If $k > 0$*, then* $0^\kappa = 0$*.*

Here are a few additional computational principles.

Proposition 18.33. *If* $\kappa \leq \lambda$ *and if* λ *is an infinite cardinal, then* $\kappa + \lambda = \lambda$*.*

Proof. If κ is a finite cardinal, then this is trivial.

We assume that κ is an infinite cardinal. Let A and B be disjoint sets such that $|A| = \kappa$ and $|B| = \lambda$. Let $\mathcal{A}$ be a collection of pairwise-disjoint collections $\{E_\alpha : \alpha \in I\}$ of subsets of B such that $|E_\alpha| = \aleph_0$ for all $\alpha \in I$. We define a partial order $\preceq$ on $\mathcal{A}$ be declaring that $\{E_\alpha : \alpha \in I\} \preceq \{F_\beta : \beta \in J\}$ if and only if $I \subseteq J$ and $E_\alpha \subseteq F_\alpha$ for all $\alpha \in I$. Each chain in $\mathcal{A}$ admits an upper bound—namely, the union—and so Zorn's lemma (Theorem 18.21) furnishes a maximal collection $\mathcal{E}_* = \{E_\alpha : \alpha \in I_*\}$ in $\mathcal{A}$.

We claim that $\mathcal{E}_*$ is a partition of B. To this end, we set

$$C = \bigcup_{\alpha \in I_*} E_\alpha.$$

and assume for a contradiction that $D = B \setminus C$ is nonempty. We note that D must be finite, for otherwise we can extract a countable subset of D and throw it into $\mathcal{E}_*$, contradicting the maximality of $\mathcal{E}_*$. Since D is finite, we see that $|E_\alpha| = |E_\alpha \cup D|$ for all $\alpha \in I_*$. Since D is nonempty, we can fix an index α_0 and replace E_{α_0} with $E_{\alpha_0} \cup D$ to obtain a larger collection than $\mathcal{E}_*$, contradicting maximality. It follows that $C = B$.

Since $|E_\alpha| = \aleph_0$ for each $\alpha \in I_*$, we see that $|B| = |\mathbb{N} \times I_*|$. Let $\mathbb{E}$ and $\mathbb{O}$ denote the sets of even natural numbers and odd natural numbers, respectively. Since $|\mathbb{N}| = |\mathbb{E}| = |\mathbb{O}|$, we see that

$$\begin{aligned} |B| &= |\mathbb{N} \times I_*| = |\mathbb{N}||I_*| \\ &= |\mathbb{E}||I_*| = |\mathbb{E} \times I_*| \\ &= |\mathbb{O}||I_*| = |\mathbb{O} \times I_*|. \end{aligned}$$

Now, $|A| \leq |B| = |\mathbb{E} \times I_*|$, and so

$$|A \cup B| = |A| + |B| \leq |\mathbb{E} \times I_*| + |\mathbb{O} \times I_*| = |\mathbb{E} \times I_* \cup \mathbb{O} \times I_*| = |\mathbb{N} \times I_*| = |B|.$$

Conversely, $|B| \leq |A \cup B|$, and so it follows from the Cantor–Bernstein theorem (Theorem 18.25) that $|A \cup B| = |B|$. □

Proposition 18.34. *If* $\kappa \leq \lambda$ *and if* λ *is an infinite cardinal, then* $\kappa \cdot \lambda = \lambda$*.*

Proof. If κ is a finite cardinal, then we see at once that $\kappa \cdot \lambda = \sum_{i=1}^{\kappa} \lambda$. Proposition 18.33 implies that $\sum_{i=1}^{\kappa} \lambda = \lambda$.

Suppose that κ is infinite. We assume for now that $\lambda = \kappa$ and establish the desired result. If $\kappa = \aleph_0$, then it suffices to note that the map $f : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ given by the formula $f(m, n) = 2^m(2n + 1) - 1$ is a bijection. We suppose that $\kappa \geq \aleph_0$ and pick a set X such that $|X| = \kappa$. We let D_0 be a countably infinite subset of X.

Set

$$\mathcal{A} = \{(a, D) : D \subseteq X \text{ and } a : D \times D \to D \text{ is a bijection}\}.$$

We have shown that $|D_0 \times D_0| = \aleph_0 \cdot \aleph_0 = \aleph_0 = |D_0|$, and so $\mathcal{A}$ is nonempty. We define a partial order $\preceq$ on $\mathcal{A}$ by declaring $(a, D) \preceq (a', D')$ if and only if $D \subseteq D'$ and $a = a'|_D$. If $\{(a_\alpha, D_\alpha) : \alpha \in I\}$ is a chain in $\mathcal{A}$, then the ordered pair (a, D_∞) defined by setting $D_\infty = \bigcup_{\alpha \in I} D_\alpha$ and

$$a(x) = a_\alpha(x) \text{ for any } \alpha \in I \text{ such that } x \in D_\alpha$$

is an upper bound of the chain. Zorn's lemma (Theorem 18.21) now furnishes a maximal pair (A, D) in $\mathcal{A}$. This, in particular, implies that $|D \times D| = |D|$.

It suffices to check that $D = X$, as this implies that

$$\kappa \cdot \kappa = |X \times X| = |X| = \kappa.$$

To this end, we suppose for a contradiction that $D \neq X$. If $|D| = |X|$, then we can find a bijection $b : X \to D$, whence we can define an extension $A' : X \times X \to X$ of A by setting $A'(x, y) = (b^{-1} \circ A)(b(x), b(y))$. As A' is a bijection, the maximality condition on (A, D) is violated.

We now assume that $|D| < |X|$ and set $E = X \smallsetminus D$. We must have $|E| = |X|$, for otherwise Proposition 18.33 implies that $X = D \cup E$ is equinumerous to either D or E, which is absurd. Let us now fix a subset E' of E such that $|E'| = |D|$. Since D and E are disjoint, we see that $E' \times E'$, $E' \times D$, and $D \times D'$ are pairwise disjoint. Since

$$\begin{aligned} |E' \times E'| &= |E'||E'| = |D||D| = |D| \\ |E' \times D| &= |E'||D| = |D||D| = |D| \\ |D \times E'| &= |D||E'| = |D||D| = |D| \end{aligned}$$

we invoke Proposition 18.33 to conclude that

$$G = (E' \times E') \cup (E' \times D) \cup (D \times E')$$

is equinumerous to D. We let $g : E' \to G$ be the corresponding bijection.

We now define a bijection $h : D \cup E' \to (D \times D) \cup G$ by setting

$$h(x) = \begin{cases} A^{-1}(x) & \text{if } x \in D; \\ g(x) & \text{if } x \in E'. \end{cases}$$

(h is well-defined, as $D \cap E' = \varnothing$.) It follows that $(h, D \cup E) \succeq (A, D)$, contradicting the maximality condition on (A, D). We therefore conclude that

$$\kappa \cdot \kappa = \kappa$$

for every infinite cardinal κ.

We now suppose that κ and λ are infinite cardinals such that $\kappa \leq \lambda$. To this end, we observe that

$$\lambda = 1 \cdot \lambda \leq \kappa \cdot \lambda.$$

By what we have proved above,

$$\kappa \cdot \lambda \leq \lambda \cdot \lambda = \lambda.$$

It now follows from the Cantor–Bernstein theorem (Theorem 18.25) that $\lambda = \kappa \cdot \lambda$. $\square$

Proposition 18.35. *Let $\{\kappa_\alpha : \alpha \in I\}$ be an indexed set of cardinal numbers. We define the* infinite sum *of the cardinals in the collection to be*

$$\sum_{\alpha \in I} \kappa_\alpha = \left| \bigcup_{\alpha \in I} X_\alpha \right|,$$

where $\{X_\alpha \in \alpha \in I\}$ is a pairwise-disjoint collection of sets such that $|X_\alpha| = \kappa_\alpha$ for all $\alpha \in I$. It is true that, if λ is a cardinal number such that $\lambda \leq |I|$ and if $\kappa_\alpha > 0$ for all $\alpha < \lambda$, then

$$\sum_{\alpha < \lambda} \kappa_\alpha = \lambda \cdot \sup_{\alpha < \lambda} \kappa_\alpha.$$

In particular, if $\kappa_\alpha = \kappa$ for all $\alpha \in I$, then

$$\sum_{\alpha \in I} \kappa = \kappa \cdot |I|.$$

Proof. If $\kappa_\alpha = \kappa$ for all α, then we can assume that there exists a set X such that $X_\alpha = X$ for all α and consider the bijection $f : I \times X \to \bigcup_{\alpha \in I} X_\alpha$ given by the formula $f(\alpha, x) = x_\alpha$, where x_α is the unique element of X_α such that $x_\alpha = x$.

We now assume that κ_α are not necessarily equal to one another. Let $\kappa_* = \sup_{\alpha < \lambda}$ and $\sigma = \sum_{\alpha < \lambda} \kappa_\alpha$. Since $\kappa_\alpha \leq \kappa$ for all α, we see that

$$\sum_{\alpha < \lambda} \kappa_\alpha \leq \sum_{\alpha < \lambda} \kappa_* = \kappa_* \cdot \lambda.$$

Conversely, $\kappa_\alpha \geq 1$ for all α, and so

$$\sum_{\alpha < \lambda} \kappa_\alpha \geq \sum_{\alpha < \lambda} 1 = \lambda.$$

Moreover, $\sigma \geq \kappa_\alpha$ for all α, and so

$$\sum_{\alpha < \lambda} \kappa_\alpha \geq \kappa_*.$$

It follows from Proposition 18.34 that

$$\sum_{\alpha < \lambda} \kappa_\alpha \geq \max(\lambda, \kappa_*) = \lambda \cdot \kappa_*.$$

The Cantor–Bernstein theorem (Theorem 18.25) now implies that $\sigma = \lambda \cdot \kappa_*$. □

19. Appendix: Analytic Miscellanies

In this appendix, we collect miscellaneous analytic facts about vector spaces relevant to the main exposition.

19.1. **Quotient Spaces of a Normed Linear Space.**

In this subsection, we develop a version of the quotient space formalism and the isomorphism theorems that come with it, as developed in Section 5, for normed linear spaces.

Definition 19.1. Let $(X, \|\cdot\|)$ be a normed linear space and Y a subspace of X. The *quotient norm* on the quotient vector space X/Y is given by

$$\|[x]\| = \inf_{z \in [x]} \|z\| = \inf_{y \in Y} \|x + y\|. \tag{19.2}$$

Note that the canonical projection $\pi : X \to X/Y$ is a bounded linear operator. Indeed,

$$\|\pi(x)\| = \inf_{z\in[x]} \|z\| \leq \|x\|.$$

The next result shows that act of taking quotients preserves completeness in many cases.

Proposition 19.3. *If X is a Banach space and Y is a closed linear subspace of X, then X/Y is a Banach space.*

Proof. Let $([x_n])_{n=1}^\infty$ be a Cauchy sequence in X/Y and extract a subsequence $([x_{n_k}])_{k=1}^\infty$ such that

$$\|[x_{n_k} - x_{n_{k+1}}]\| = \|[x_{n_k}] - [x_{n_{k+1}}]\| < 2^{-k}$$

for each $k \in \mathbb{N}$. We now pick $z_0 \in [x_1]$ and fix, for each $k \in \mathbb{N}$, an element $z_k \in [x_{n_k} - x_{n_{k+1}}]$ such that $\|z_k\| < 2^{-k}$. Setting $a_K = z_0 + \cdots + z_K$, we see that $(a_k)_{k=1}^\infty$ is a Cauchy sequence in X, whence it converges to a limit $a \in X$. Now,

$$a - a_K = \sum_{k=K+1}^\infty z_k$$

is an element of $[a] - [a_K] = [a] - [x_{n_K}] = [a - x_{n_K}]$, and so

$$\|[a] - [x_{n_K}]\| \leq \left\| \sum_{k=K+1}^\infty z_k \right\| \leq \sum_{k=K+1}^\infty \|z_k\| < 2^{-K},$$

It follows that $[x_{n_K}] \to [a]$ in X/Y, and the desired result follows. □

The first isomorphism theorem generalizes readily to the category of Banach spaces, taken as a subcategory of the category of topological vector spaces. We remark that the isomorphism we get is *not* an isometric isomophism in general.

Theorem 19.4 (First isomorphism theorem for Banach spaces)**.** *If $T : X \to Y$ is a bouned linear transformation, then*

$$X/\ker T \cong \operatorname{im} T$$

in the category of topological vector spaces.

Proof. We take the linear isomorphism $\tilde{T} : X/\ker T \to T$ constructed in the proof of Theorem 5.11. We observe that

$$\|\tilde{T}[x]\| = \|Tz\| \leq \|T\|\|z\|$$

for each $z \in [x]$, whence taking the infimum over $[x]$ yields

$$\|\tilde{T}[x]\| \leq \|T\|\|[x]\|.$$

Therefore, $\tilde{T}$ is bounded. It now follows from the open mapping theorem (Corollary 10.6) that $\tilde{T}^{-1}$ is bounded. □

19.2. **Annihilators.** In this subsection, we generalize the notion of orthogonal complements to Banach spaces.

Definition 19.5. Let X be a topological vector space. The *annihilator* of a subset Y of X is the collection of all continuous linear functionals $l : X \to \mathbb{F}$ that vanish on Y. The annihilator of Y is denoted by $Y^\perp$.

If X is a Hilbert space, then the Riesz representation theorem (Theorem 16.26) implies that $Y^\perp$ is precisely the orthogonal complement (Definition 16.20) of Y.

If X is a normed linear space and Y a closed subspace of X, then Corollary 11.13 guarantees that $Y^\perp$ is nontrivial as long as Y is a proper subspace. The following proposition quantifies this phenomenon precisely.

Proposition 19.6. *If X is a Banach space and Y a closed linear subspace of X, then $Y^\perp \cong (X/Y)^*$ and $Y^* \cong X^*/Y^\perp$ in the category of topological vector spaces.*

Proof. To establish the first isomorphism, we let $\pi : X \to X/Y$ be the canonical projection onto the quotient space X/Y. Observe that the adjoint $\pi^* : (X/Y)^* \to X^*$ is injective. Indeed, if $\pi^*(l_1) = \pi^*(l_2)$, then

$$l_1 = l_1\pi = \pi^*(l_1) = \pi^*(l_2) = l_2\pi = l_2$$

by the surjectivity of π. In particular, $\ker \pi^*$ is trivial, and the first isomorphism theorem (Theorem 19.4) implies that $(X/Y)^* \cong Y^\perp$, provided that $\operatorname{im} \pi^* = Y^\perp$. But this is trivial, as $\pi^*(l) = l\pi$ are precisely the linera functionals that vanish on Y.

To establish the second isomorphism, we let $\iota : Y \to X$ be the canonical embedding. Observe that the adjoint $\iota^* : X^* \to Y^*$ is surjective. Indeed, if $l \in Y^*$, then the analytic Hahn-Banach theorem (Theorem 11.12) furnishes an $L \in X^*$ such that

$$\iota^*(L) = L\iota = L|_Y = l.$$

It now follows from the first isomorphism theorem (Theorem 19.4) that implies that $X^*/Y^\perp \cong Y^*$ in the category of topological vector spaces. provided that $\ker \iota^* = Y^\perp$. But this is trivial, as $\iota^*(l) = l\iota = l|_Y$. □

We remark that the isomorphism $Y^\perp \cong (X/Y)^*$ by itself does not guarantee the nontriviality of $Y^\perp$. The Hahn-Banach theorem, however, guarantees that X^* is always "bigger" than Y^*. Therefore, if Y is a proper subspace, then $Y^\perp$ must be nontrivial to make up for this difference, as quantified by the isomorphism $Y^* \cong X^*/Y^\perp$.

Tacitly used in the proof of the above proposition is an assumption that annihilators are subspaces. In fact, the quotient construction requires $Y^\perp$ to be a *closed* subspace of X^*. We address this in the following proposition, which does not require completeness.

Proposition 19.7. *If Y is a subset of a normed linear space X, then $Y^\perp$ is a closed linear subspace of X^*. Furthermore,*

$$Y^\perp = [\overline{\operatorname{span}(Y)}]^\perp. \tag{19.8}$$

Proof. Whenever $l_1, l_2 \in Y^\perp$ and $\lambda \in \mathbb{F}$

$$(\lambda l_1 + l_2)(y) = \lambda l_1(y) + l_2(y) = 0 + 0 = 0$$

for all $y \in Y$. Therefore, $Y^\perp$ is a linear subspace of X^*. To show that $Y^\perp$ is closed, we pick a sequence $(l_n)_{n=1}^\infty$ in $Y^\perp$ converging to $l \in X^*$ and observe that

$$|l(y)| = |l(y) - l_n(y)| = \lim_{n\to\infty} |l(y) - l_n(y)| \leq \lim_{n\to\infty} \|l - l_n\|\|y\| = 0$$

for all $y \in Y$.

It remains to verify the formula (19.8). $Y \subseteq \overline{\text{span}(Y)}$ clearly implies that $Y^\perp \supseteq [\overline{\text{span}(Y)}]^\perp$, and so it suffices to show the reverse inclusion. But for each $l \in Y^\perp$, $l|_{\text{span}(Y)} = 0$ holds trivially and the desired result now follows form the continuity of l. ☐

Experiences with finite-dimensional inner-product spaces suggest that the orthogonal complement of the orthogonal complement of a subspace should be the subspace itself. This is not true in general, but we can prove an analogous result for reflexive Banach spaces.

Proposition 19.9. *If Y is a subset of a reflexive Banach space X, then*

$$(Y^\perp)^\perp \cong \overline{\text{span}(Y)}.$$

In particular, if Y is a closed linear subspace, then $(Y^\perp)^\perp \cong Y$.

Proof. We prove the "in particular" part of the proposition, from which we can derive the proposition in its full generality via Proposition 19.7. We therefore assume that Y is a closed linear subspace of X. By Proposition 19.6, we see that

$$(Y^\perp)^\perp \cong (X^*/Y^\perp)^* \cong (Y^*)^*.$$

Because every closed subspace of a reflexive Banach space is reflexive[9], it follows that $(Y^\perp)^\perp \cong Y$. ☐

Annihilators provide a useful way of thinking about dense subspaces of a space whose dual space admits a concrete characterizations.

Example 19.10. Let $1 \leq p < \infty$. By the L^p Riesz representation theorem, the dual of $L^p(X,\mu)$ is isometrically isomorphic to $L^q(X,\mu)$, where $1/p + 1/q = 1$. By Propositions 19.6 and 19.7, a subspace $\mathcal{D}$ of $L^p(X,\mu)$ is dense in $L^p(X,\mu)$ if and only if each $u \in L^q(X,\mu)$ satisfying

$$\int_X fu \, d\mu = 0$$

for all $f \in \mathcal{D}$ must be the zero function. Standard examples of dense subspaces of $L^p(\mathbb{R}^d)$ include the space of simple functions on cubes, the space $\mathscr{C}(\mathbb{R}^d)$ of continuous functions, the Schwartz space $\mathscr{S}(\mathbb{R}^d)$, and the space $\mathscr{C}_c^\infty(\mathbb{R}^d)$ of continuous functions with compact support.

Familiar examples of denses subspaces abound if we restrict our attention to compact subsets of $\mathbb{R}^d$. For example, the Weierstrass approximation theorem asserts that polynomials on $[a,b]$ are dense in $\mathscr{C}([a,b])$, which is then dense in $L^p([a,b])$. It now follows from (19.8) in Proposition 19.7 that any $u \in L^q$ satisfying $\int_a^b f(x)x^n \, dx = 0$ for all $n \in \mathbb{N}$ must be the zero function. Similarly, the L^2 Fourier inversion theorem shows that trigonometric polynomials on $[-\pi,\pi]$ are dense in $L^2([-\pi,\pi])$, whence any $f \in L^2$ satisfying $\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-inx}\, dx$ for all $n \in \mathbb{N}$

[9]See [Bre11], Proposition 3.20 for a proof via Kakutani's theorem; see [Lax02], Chapter 8, Theorem 15 for a proof without Kakutani.

must be the zero function. In particular, the Fourier series of an L^2-function is unique.

19.3. **Moore–Smith Theory of Nets.** In this subsection, we develop the theory of nets, a topological generalization of the theory of sequences. of nets. In particular, we focus on generalizing the sequential characterization of compactness. Recall that sequential compactness and compactness are equivalent on metric spaces.

Theorem 19.11 (Sequential criterion for compactness in metric spaces)**.** *A set K in a metric space is compact if and only if every sequence in K admits a convergent subsequence.*

Unfortunately, this criterion does not hold in the category of topological spaces.

Proposition 19.12. *The space $[0,1]^{[0,1]}$ of $[0,1]$-fold product of $[0,1]$ with the product topology is compact but not sequentially compact.*

Proof. By Tychonoff's theorem (Theorem 19.25), $[0,1]^{[0,1]}$ is compact. To see that $[0,1]^{[0,1]}$ is not sequentially compact, we define the nth binary expansion function $B_n : [0,1] \to [0,1]$ by setting $B_n(x)$ to be the nth digit of the binary expansion of x, so that $(B_n)_{n=1}^\infty$ is a sequence in $[0,1]^{[0,1]}$. Given a subsequence $(B_{n_k})_{k=1}^\infty$, we fix $x \in [0,1]$ such that

$$B_{n_k}(x) = \begin{cases} 0 & \text{if } k \text{ is odd;} \\ 1 & \text{if } k \text{ is even.} \end{cases}$$

Then $(B_{n_k}(x))_{k=1}^\infty$ clearly does not converge, whence neither does $(B_{n_k})_{k=1}^\infty$ in $[0,1]^{[0,1]}$. □

Proposition 19.13. *The first uncountable ordinal ω_1 (§18.2) with the order topology is sequentially compact but not compact.*

Proof. Observe first that ω_1 is noncompact, for the open cover

$$\{[0,\kappa) : \kappa \in [0,\omega_1)\}$$

clearly doesn't admit a finite subcover.

On the other hand, ω_1 is sequentially compact. To see this, we pick an arbitrary sequence $(\kappa_n)_{n=1}^\infty$ of ordinals in ω_1. We define the *limit superior* of the sequence by setting

$$\limsup_n \kappa_m = \min_{n \geq 0} \sup_{m \geq n} \kappa_m,$$

which exists by the well-ordering of ordinals. Note that

$$\limsup_n \kappa_m = \min_{n \geq 0} \bigcup_{m \geq n} \kappa_m.$$

Since a countable union of countable ordinals is a countable ordinal, we see that $\kappa = \limsup_n \kappa_m$ is a countable ordinal.

It therefore suffices to produce a subsequence subsequence of (κ_n) that converges to κ, which we could if κ were a limit point. We suppose for a contradiction that there exists an open interval (μ,ν) containing κ that does not contain any term of the sequence (κ_n). This, in particular, implies that κ cannot be attained as the supremum of a subsequence. By the well-ordering of ordinals, however, $\kappa = \sup_{m \geq N} \kappa_m$ for some $N \in \mathbb{N}$, which is absurd. □

Similarly, it is not necesarily true that $f(x_n) \to f(x)$ for all convergent sequences (x_n) implies the continuity of f. Sequential characterizations provide a convenient way to think about topological properties concretely, so it is desirable to introduce the notion of generalized sequences that will allow us to recover them. Let us first introduce generalized index sets:

Definition 19.14. A *directed set* is a partially ordered set $(D, \leq)$ such that every pair of elements α and β of D has a common maximum γ in D, i.e., $\gamma \geq \alpha$ and $\gamma \geq \beta$.

Note that the set of natural numbers with the usual ordering is a directed set. Observing that sequences are functions on the directed set $\mathbb{N}$, we make the following generalization:

Definition 19.15. A *net* in a topological space X is a function $\Phi : D \to X$ on a directed set D.

For convenience, we typically write $(x_\beta)_{\beta \in D}$ or (x_β) to detnote a net. The convergence of a net is defined analogously to that of a sequence:

Definition 19.16. A net (x_β) in X is *eventually in* $U \subseteq X$ if there exists an index β_0 such that $x_\beta \in U$ for all $\beta \geq \beta_0$. (x_β) *converges to* $x \in X$ if, for each neighborhood U of x, the net (x_β) is eventually in U.

Immediately, we recover the sequential characterization of continuity.

Proposition 19.17. *A function $f : X \to Y$ between topological spaces is continuous if and only if, for each $x \in X$ and every net (x_β) in X converging to x, the net $(f(x_\beta))$ converges to $f(x)$.*

Defining subnets requires some care. We would like to recover, for example, the Bolzano-Weierstrass property, but the notion of subnets given by truncating the initial terms of a net is too restrictive for this.

Definition 19.18. A function $h : D' \to D$ between directed sets is *final* if, for each $\beta_0 \in D$, there exists $\beta_0' \in D'$ such that $\beta' \geq \beta_0'$ implies $h(\beta') \geq \beta_0$. A *subnet* of a net $\Phi : D \to X$ in X is the composite map $\Phi \circ h$ with a final function h.

At this point, we present the required machinery for the net-theoretic proof of Tychonoff's theorem. The key concept, whose clever definition absorbs the bulk of the work, is as follows:

Definition 19.19. A net (x_β) in X is *universal* if, for each subset A of X, the net (x_β) is eventually in either A or $X \smallsetminus A$.

The next lemma, due to Kelley, shows that universal nets exist. In fact, there are lots of them. For the sake of brevity, we introduce the following definition.

Definition 19.20. A net (x_β) is *frequently in* $A \subseteq X$ if, for each index β_0, there exists an index $\beta \geq \beta_0$ such that $x_\beta \in A$.

Note that (x_β) is frequently in A if and only if (x_β) is *not* eventually in $X \smallsetminus A$.

Lemma 19.21 (Kelley's theorem)**.** *Every net has a universal subnet.*

Proof. Let $(x_\beta)_{\beta \in D}$ be a net in a topological space X. We consider a collection $\mathscr{D}$ of all subsets $\mathcal{C}$ of $\mathcal{P}(X)$ such that

(1) if $A \in \mathcal{C}$, then (x_β) is frequently in A, and
(2) if $A, B \in \mathcal{C}$, then $A \cap B \in \mathcal{C}$.

$\mathscr{D}$ is nonempty, as $\mathcal{C} = \{X\}$ is such a set. The inclusion relation is a partial order on $\mathscr{D}$, and any chain on $\mathscr{D}$ has a maximum—namely, the union of the sets in the chain. It follows from Zorn's lemma that $(\mathscr{D}, \subseteq)$ has a maximal element $\mathcal{C}_0$.

We set $D' = \{(A, \beta) \in \mathcal{C}_0 \times D : x_\beta \in A\}$ and define a partial order $\leq$ on D' by setting

$$(A_1, \beta_1) \leq (A_2, \beta_2) \quad \text{if and only if} \quad A_1 \supseteq A_2 \text{ and } \beta_1 \leq \beta_2.$$

If $(A_1, \beta_1), (A_2, \beta_2) \in D'$, then $A = A_1 \cap A_2$ is in $\mathcal{C}_0$. Moreover, β_1 and β_2 have a common maximum β', and we can find $\beta \geq \beta'$ such that $x_\beta \in A$. It follows that (A, β) is a common maximum of (A_1, β_1) and (A_2, β_2), whence D' is a directed set. The function $(A, \beta) \mapsto \beta$ is a final map from D' to D, hence we obtain a subnet of (x_β), which we denote by $(x_{(A,\beta)})_{(A,\beta) \in D'}$.

We shall show that $(x_{(A,\beta)})$ is universal. We fix a subset E of X and suppose that $(x_{(A,\beta)})$ is frequently in E. We shall show that $(x_{(A,\beta)})$ is eventually in E, which establishes the desired result. For each index (A_0, β_0), we can find another index $(A, \beta) \geq (A_0, \beta_0)$ such that $x_{(A,\beta)} \in E$. By definition, $x_\beta = x_{(A,\beta)}$ is an element of A, whence $x_\beta \in E \cap A$. Furthermore, $A_0 \supseteq A$, and so $E \cap A_0 \supseteq E \cap A$. It follows that $x_\beta \in E \cap A_0$. Since β_0 was arbitrary, we conclude that (x_β) is frequently in $E \cap A_0$. By maximality of $\mathcal{C}_0$, the collection $\mathcal{C}_0$ must include E and $E \cap A_0$ for all $A_0 \in \mathcal{C}_0$. In particular, $(x_{(A,\beta)})$ cannot be frequently in $X \smallsetminus E$, for otherwise $E \cap (X \smallsetminus E) = \varnothing$ is in $\mathcal{C}_0$, which is absurd. We conclude that $(x_{(A,\beta)})$ is not frequently in $X \smallsetminus E$, hence eventually in E. □

We now establish three useful characterizations of compactness.

Lemma 19.22. *Let X be a topological space. The following are equivalent.*

(1) X is compact.
(2) ***Universal net criterion****. Every universal net in X converges.*
(3) ***Bolzano-Weierstrass property****. Every net in X has a convergent subnet.*
(4) ***Finite intersection property criterion****. If a collection of closed sets in X has the* finite intersection property*—the intersection of any finite subcollection is nonempty—then the intersection of the entire collection is nonempty.*

Proof. (1) ⇒ (2). Let X be a compact space and suppose for a contradiction that there exists a universal net (x_β) that does not converge. For each $x \in X$, we pick a neighborhood U_x such that (x_β) is *not* eventually in U_x. By universality, (x_β) is eventually in $X \smallsetminus U_x$, whence we can find an index β_x such that $\beta \geq \beta_x$ implies $x_\beta \notin U_x$. We now extract a finite subcover $\{U_{x_1}, \ldots, U_{x_n}\}$ of the open cover $\{U_x\}_{x \in x}$ and set $\beta_0 = \max\{\beta_{x_1}, \ldots, \beta_{x_n}\}$. Then, for each $\beta \geq \beta_0$, we see that $x_\beta \notin U_{x_i}$ for all $1 \leq i \leq n$, whence $x \notin X$. This is evidently absurd, and we conclude that every universal net converges.

(2) ⇒ (3). This is a trivial consequence of Kelley's theorem (Lemma 19.21).

(3) ⇒ (4). We let $\mathscr{C}$ be a collection of closed subsets of X with the finite intersection property. Adding in finite intersections of elements of $\mathscr{C}$ does not break the finite intersection property, so we assume without loss of generality that $\mathscr{C}$ is closed under finite intersection. This turns $\mathscr{C}$ into a directed set with $\supseteq$ as the partial order. We define a net on $\mathscr{C}$ by fixing an element x_C of each index set $C \in \mathscr{C}$.

By the Bolzano-Weiestrass property, we can find a directed set D and a final map $f : D \to \mathscr{C}$ such that the subnet $(x_{f(\beta)})_{\beta \in X}$ converges, say, to x. For each $C \in \mathscr{C}$, we can find $\beta_C \in D$ such that $\beta \geq \beta_C$ implies $f(\beta) \geq C$. Unraveling the definitions, we have the inclusion relation

$$x_{f(\beta)} \in f(\beta) \subseteq C$$

for all $\beta \geq \beta_C$, whence by the closedness of C the point x must be in C. Since C was arbitrary, it follows that $x \in \bigcap_{C \in \mathscr{C}} C$, as was to be shown.

(4) $\Rightarrow$ (1) Let $\{U_\alpha\}$ be an open cover of X. Let $C_\alpha = X \smallsetminus U_\alpha$ for each index α and consider the collection $\mathscr{C} = \{C_\alpha\}$. Since $\bigcup U_\alpha = X$, the intersection $\bigcap C_\alpha$ is empty, whence the contrapositive of (4) implies that there exists at least one finite subcollection $\{C_1, \ldots, C_n\}$ of $\mathscr{C}$ whose intersection is empty. Setting $U_n = C_n$, this implies that the union of the subcollection $\{U_1, \ldots, U_n\}$ of $\{U_\alpha\}$ is X, whence $\{U_1, \ldots, U_n\}$ is the desired finite subcover. $\square$

We are almost ready to present a proof of Tychonoff's theorem. We require two more relatively simple lemmas.

Lemma 19.23. *If $\Phi : D \to X$ is a universal net and $f : X \to Y$ an arbitrary function between topological spaces, then $f \circ \Phi$ is a universal net in Y.*

Proof. We let $x_\beta = \Phi(\beta)$ for notational convenience. Given $A \subseteq Y$, the net (x_β) is eventually in either $f^{-1}(A)$ or $X \smallsetminus f^{-1}(A) = f^{-1}(Y \smallsetminus A)$. It follows that $f(x_\beta)$ is in either A or $Y \smallsetminus A$. $\square$

Lemma 19.24. *A net $(x_\beta)_\beta$ in $\prod_\alpha X_\alpha$ converges to $x = (x^\alpha)$ if and only if each component net $(x^\alpha_\beta) = (\pi_\alpha(x_\beta))_\beta$ converges to $x^\alpha = \pi_\alpha(x)$.*

Proof. Suppose that $x_\beta \to x$. If U_α is a neighborhood of x^α, then we can find an index β_0 such that $x_\beta \in \pi^{-1}(U_\alpha)$ for all $\beta \geq \beta_0$, whence $x^\alpha_\beta \in U_\alpha$ for all $\beta \geq \beta_0$. It follows that $x^\alpha_\beta \to x^\alpha$.

Conversely, we suppose that $x^\alpha_\beta \to x^\alpha$ for all α. If U is a neighborhood of x, then we can find neighborhoods U_α of x^α such that $\prod_\alpha U_\alpha \subseteq U$ and $U_\alpha = X$ for all but finitely many α. If $U_\alpha = X_\alpha$, then x^α_β is trivially in U_α. We let $\alpha_1, \ldots, \alpha_n$ be the indices such that $U_{\alpha_i} \neq X_\alpha$, and find an index β_i such that $\beta \geq \beta_i$ implies $x^{\alpha_i}_\beta \in U_{\alpha_i}$. Setting $\beta_0 = \max\{\beta_1, \ldots, \beta_n\}$, we see that $\beta \geq \beta_0$ implies $x^\alpha_\beta \in U_\alpha$ for all α, hence in U. It follows that $x_\beta \to x$. $\square$

Finally, the theorem and the proof:

Theorem 19.25 (Tychonoff)**.** *If $\{X_\alpha\}$ is a family of compact topological spaces, then the product topology on $\prod_\alpha X_\alpha$ turns $\prod_\alpha X_\alpha$ into a compact topological space.*

Proof. Let (x_β) be a universal net in $\prod_\alpha X_\alpha$. By Lemma 19.23, $(x^\alpha_\beta)_\beta$ is a universal net in X_α for each α, whence by Lemma 19.22 it is a convergent net in X_α. We invoke the Axiom of Choice to fix a limit x^α of $(x^\beta_\alpha)_\beta$ for each α. By Lemma 19.24, the net (x_β) converges to $x = (x^\alpha)$ in $\prod_\alpha X_\alpha$. Therefore, every universal net in $\prod_\alpha X_\alpha$ converges, and it follows from Lemma 19.22 that $\prod_\alpha X_\alpha$ is compact. $\square$

References

[Arv76] William Arveson, *An invitation to c*-algebras*, Springer, 1976.

[Bre11] Haïm Brezis, *Functional analysis, sobolev spaces and partial differential equations*, Springer, 2011.

[Dou98] Ronald G. Douglas, *Banach algebra techniques in operator theory*, Springer, 1998.

[DR50] A. Dvoretzky and C. A. Rogers, *Absolute and unconditional convergence in normed linear spaces*, Proceedings of the National Academy of Sciences **36** (1950), no. 3, 192–197.

[GN43] Israel Gelfand and Mark Naimark, *On the embedding of normed rings into the ring of operators in hilbert space*, Mathematicheskii Sbornik **12** (1943), no. 2, 197–217.

[Gra10a] Loukas Grafakos, *Classical fourier analysis*, Springer, 2010.

[Gra10b] ______, *Modern fourier analysis*, Springer, 2010.

[HJ99] Karel Hrbacek and Thomas Jech, *Introduction to set theory*, CRC Press, 1999.

[Jec06] Thomas Jech, *Set theory*, third ed., Springer, 2006.

[Lax02] Peter D. Lax, *Functional analysis*, Wiley-Interscience, 2002.

[Lec] *David Lecomte's lecture notes on weak topologies*, http://perso.crans.org/lecomte/Math/WeakTopologies.pdf.

[Mac98] Saunders Mac Lane, *Categories for the working mathematician*, second ed., Springer, New York, NJ, 1998.

[MO1] *When does Cantor-Bernstein hold?*, http://mathoverflow.net/questions/1058/.

[MS13] Camil Muscalu and Wilhelm Schlag, *Classical and multilinear harmonic analysis*, Cambridge University Press, 2013.

[Ric45] Charles E. Rickart, *Banach algebras with an adjoint operation*, Annals of Mathematics **47** (1945), no. 3, 528–550.

[Rud91] Walter Rudin, *Functional analysis*, second ed., McGraw-Hill, 1991.

[Seg47] Irving E. Segal, *Irreducible representations of operator algebras*, Bulletin of the American Mathematical Society **53** (1947), 73–88.

[SS11] Elias M. Stein and Rami Shakarchi, *Functional analysis: Introduction to further topics in analysis*, Princeton University Press, 2011.

[Tao10] Terence Tao, *An epsilon of room, i: Real analysis*, American Mathematical Society, 2010.

[Tes09] Gerald Teschel, *Mathematical methods in quantum mechanics: With applications to schrödinger operators*, AMerican Mathematical Society, 2009.

[Thi06] Christoph Thiele, *Wave packet analysis*, American Mathematical Society, 2006.